Chapter 6

SAMPLING DISTRIBUTIONS

6.1 Only if the pieces are put on the assembly line such that every 20-th piece is random with respect to the characteristic being measured.

If twenty molds are dumping their contents, in sequential order, onto the assembly line then the sample would consist of output from a single mold . It would not be random.

6.2 The subscribers to Literary Digest were not representative of the people voting in the 1932 Presidential election. At that time, people with telephones were also likely to have higher incomes than the typical voter. Instead, the poll over-represented upper income voters who were likely to vote Republican.

6.3 (a) A typical member of the population does not vacation on a luxury cruise. The sample taken would be biased.

(b) This sample will very likely be biased. Those will high incomes will tend to respond while those with low incomes will tend to not respond.

(c) Everyone feels that unfair things should be stopped. The way the question is phrased biases the responses.

6.4 Number the states from 1 to 50 according to their alphabetical order. Using the first two columns of Table 7 to select 8 numbers between 1 and 50 by discarding numbers outside of this range as well as any previously drawn number gives

$$13,\ 4,\ 29,\ 14,\ 36,\ 48,\ 22,\ 35.$$

These correspond to Illinois, Arkansas, New Hampshire, Indiana, Oklahoma, West Virginia, Michigan, and Ohio.

6.5 (a) The number of samples (given that order does not matter) is

$$\binom{6}{2} = \frac{6 \cdot 5}{2 \cdot 1} = 15.$$

(b) The number of samples (given that order does not matter) is

$$\binom{25}{2} = \frac{25 \cdot 24}{2 \cdot 1} = 300.$$

6.6 For a random sample, each outcome must have equal probability. Thus, the probabilities are:

(a) $\frac{1}{15} = .0667.$

(b) $\frac{1}{300} = .00333.$

6.7 (a) The probability of each of the numbers is given in the table:

Number	−4	−3	−2	−1	0	1	2	3	4
Probability	1/6	2/15	1/10	1/15	1/15	1/15	1/10	2/15	1/6

The mean of the distribution is

$$\begin{aligned}&\frac{1}{6}(-4) + \frac{2}{15}(-3) + \cdots + \frac{2}{15}(3) + \frac{1}{6}(4)\\ &\quad = \frac{1}{6}(4-4) + \frac{2}{15}(3-3) + \frac{1}{10}(2-2) + \frac{1}{15}(1-1) + \frac{1}{15}(0) = 0.\end{aligned}$$

The variance is

$$\frac{2}{6}(16) + \frac{4}{15}(9) + \frac{2}{10}(4) + \frac{2}{15}(1) + \frac{1}{15}(0) = 8.667.$$

(b) The 50 samples are shown in Table 6.1 along with their means.

Table 6.1. 50 Samples of Size 10 Taken Without Replacement

obs.	obs.	obs.	obs.	obs.	obs.	obs.	obs.	obs.	obs.	mean
1	3	-1	4	4	4	-4	-4	-4	-2	0.1
-2	3	3	-2	2	-4	4	-4	-3	-4	-0.7
4	4	-2	-3	-3	2	1	3	4	0	1.0
3	3	-3	4	2	1	-2	0	4	0	1.2
4	0	-1	-2	3	1	2	4	-1	1	1.1
1	-2	-3	-4	2	4	-2	3	4	-4	-0.1
-1	2	1	-1	-4	-3	0	-4	3	1	-0.6
-4	-2	2	1	-4	-3	3	2	-2	1	-0.6
1	-3	2	-4	2	4	3	-4	-3	-2	-0.4
1	4	4	4	-4	3	-1	-2	2	-4	0.7
4	0	-1	1	-3	1	4	-3	4	2	0.9
-4	4	2	-4	2	-2	-2	-1	4	3	0.2
2	-1	-2	4	-4	0	-4	1	3	-2	-0.3
-4	-1	-3	4	0	-4	-3	1	2	4	-0.4
-2	4	4	-1	-4	1	4	1	-4	-3	0.0
-3	0	-3	-2	0	-4	4	-1	1	-2	-1.0
4	-4	3	4	-4	-1	3	2	0	1	0.8
3	3	-3	4	-3	1	3	-3	-4	-1	0.0
3	-4	-4	4	4	3	4	-3	2	-2	0.7
4	0	2	-3	-3	-3	-2	3	-4	-4	-1.0
2	-1	4	-3	2	3	4	4	-4	4	1.5
3	2	3	4	4	4	3	1	2	-2	2.4
4	-3	0	-1	-3	-3	0	-4	4	-4	-1.0
-3	3	-4	3	0	1	-4	-4	-3	-4	-1.5
-2	-4	-2	-3	4	4	3	-3	1	-3	-0.5
3	-1	-2	-3	1	-4	-2	3	4	2	0.1
4	2	-4	3	-2	2	-2	1	4	3	1.1
0	-2	4	-2	0	-1	2	3	-3	-4	-0.3

Table 6.1.(continued)

obs.	obs.	obs.	obs.	obs.	obs.	obs.	obs.	obs.	obs.	mean
1	-4	4	4	-2	-3	-4	-4	0	3	-0.5
2	4	3	-2	3	1	0	0	-4	-3	0.4
4	0	2	-1	1	-4	-4	-4	-3	4	-0.5
4	-3	-1	-3	-3	0	4	4	3	1	0.6
-4	-1	0	-3	4	-3	2	1	3	4	0.3
0	4	-4	3	-2	2	0	4	2	-4	0.5
3	2	-2	4	-4	4	3	4	0	-3	1.1
-4	2	0	-2	-3	2	-1	-2	-4	-2	-1.4
-3	-1	-2	4	2	4	0	1	2	2	0.9
2	-4	-2	-3	-1	3	1	-2	-4	-3	-1.3
-4	3	1	0	-1	0	-3	-3	2	4	-0.1
2	2	4	-4	-4	4	4	-3	-3	-2	0.0
-4	1	4	4	3	2	-2	-3	4	-3	0.6
0	-4	-4	0	-4	1	4	-3	1	4	-0.5
-4	4	-3	1	4	-1	-4	-3	-4	-2	-1.2
4	-3	-3	4	4	-3	1	2	3	-1	0.8
-2	3	2	-4	-1	3	-4	-1	-3	1	-0.6
4	2	4	-3	-4	-4	-1	-3	-4	-2	-1.1
-3	-1	1	2	-3	0	-4	-2	4	2	-0.4
0	-1	4	-4	2	-2	3	-3	-2	3	0.0
-4	2	1	-2	-2	-1	-3	-4	3	-3	-1.3
-4	4	4	4	-3	2	3	3	3	4	2.0

(c) The mean of the 50 sample means is .034. The sample variance of the 50 sample means is .8096.

(d) According to Theorem 6.1, the distribution of the means has mean 0 and variance

$$\frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} = \frac{8.667}{10} \cdot \frac{20}{29} = .5977.$$

The sample values in part (c) compare well with these theoretical values.

6.8 (a) The mean and variance of the distribution are the same as in part (a) of Exercise 6.7.

(b) The 50 samples, taken with replacement, are shown in Table 6.2 along with their means.

Table 6.2. 50 Samples of Size 10 Taken with Replacement.

obs.	obs.	obs.	obs.	obs.	obs.	obs.	obs.	obs.	obs.	mean
-3	-4	2	1	0	-3	-1	3	4	-4	-0.5
0	1	-4	-4	0	2	4	3	2	-4	0.0
-3	-1	-3	2	-4	-3	0	4	4	-4	-0.8
4	-3	2	3	4	4	-2	2	-2	-1	1.1
2	3	-1	-3	4	-3	4	-2	-3	3	0.4
4	-2	-4	3	-4	3	2	1	-4	-2	-0.3
0	3	4	4	0	4	0	2	-2	2	1.7
4	-2	-3	-2	3	-4	-3	2	4	0	-0.1
-4	1	-3	4	-3	4	-3	3	4	-2	0.1
-1	-4	-3	4	-4	0	-4	2	4	4	-0.2
4	-3	-4	-2	4	-2	-4	-1	-4	2	-1.0
4	4	3	4	4	-4	-2	-1	-1	-4	0.7
4	3	3	2	3	-2	-4	4	-4	-4	0.5
4	-3	2	4	4	1	-2	3	-1	1	1.3
-1	-2	4	-4	4	-3	-2	-2	-3	-4	-1.3
-3	3	0	-4	4	-4	3	-2	3	4	0.4
-1	-4	0	3	-4	0	3	-3	4	-3	-0.5
3	4	2	-1	-3	0	-3	2	3	-3	0.4
-4	3	-4	-3	-4	-4	-2	-3	1	-4	-2.4
-4	-2	3	-3	-3	4	3	4	-3	-3	-0.4
1	4	1	-4	-2	0	3	3	3	2	1.1
-2	4	2	4	0	2	-4	-1	4	3	1.2
-4	-1	-1	-3	-3	-2	-2	-3	2	-2	-1.9
4	3	3	-1	0	-4	-2	4	0	-3	0.4
-3	1	0	-4	3	3	2	3	-3	-1	0.1
-2	0	2	-2	4	-4	4	-4	2	-4	-0.4
-2	-4	-3	0	-3	4	-3	4	-3	3	-0.7
-4	-1	2	4	-3	-3	2	1	4	-4	-0.2
2	-2	-4	-4	-4	0	-1	2	-4	-4	-1.9
-3	-2	0	-4	-1	1	-4	1	0	3	-0.9
-2	-2	1	3	4	-2	4	2	-4	4	0.8
3	1	-3	4	2	4	-2	-2	-4	-4	-0.1

Table 6.2.(continued)

obs.	obs.	obs.	obs.	obs.	obs.	obs.	obs.	obs.	obs.	mean
2	-2	-4	2	1	2	4	-2	2	1	0.6
4	-2	-3	3	4	-2	1	3	4	-1	1.1
-4	4	4	-4	-4	3	-3	4	3	-3	0.0
-2	0	3	-4	1	2	2	1	-2	-1	0.0
-2	4	4	4	3	-2	1	-2	-2	3	1.1
-4	-2	1	3	-4	-2	-4	-1	4	0	-0.9
-3	0	3	4	2	-3	1	-2	4	1	0.7
-4	0	-1	4	3	-1	-2	4	3	2	0.8
-4	4	4	0	-3	0	3	4	3	3	1.4
-4	-3	-3	3	2	1	-4	1	-2	-2	-1.1
3	-4	-2	0	3	0	-3	-3	-3	3	-0.6
-3	3	-2	4	-1	2	-1	1	2	4	0.9
-2	-3	-2	4	-1	-4	-3	3	-3	4	-0.7
-3	3	4	3	-2	-1	-1	0	3	-4	0.2
-4	-3	-4	4	1	-4	0	-3	4	-4	-1.3
-3	-4	-4	0	-4	4	0	2	4	-4	-0.9
-3	-3	-2	-4	4	4	4	-2	-4	-3	-0.9
3	3	-2	1	-3	0	2	3	-1	0	0.6

(c) The mean of the means is $-.048$. The sample variance of the means is .8621.

(d) According to Theorem 6.1, the distribution of the means has mean 0 and variance

$$\sigma^2/n = 8.667/10 = .8667.$$

The sample values in part (c) compare well with these theoretical values.

6.9 A table of each outcome and the mean follows:

Outcome	Mean	Outcome	Mean
1, 1	1.0	3, 1	2.0
1, 2	1.5	3, 2	2.5
1, 3	2.0	3, 3	3.0
1, 4	2.5	3, 4	3.5
2, 1	1.5	4, 1	2.5
2, 2	2.0	4, 2	3.0
2, 3	2.5	4, 3	3.5
2, 4	3.0	4, 4	4.0

Thus , the distribution of these values is:

Value	No. of Ways Obtained	Probability
1.0	1	.0625
1.5	2	.1250
2.0	3	.1875
2.5	4	.2500
3.0	3	.1875
3.5	2	.1250
4.0	1	.0625

where the probability is (*no. ways*)$\times(.25)^2$. Consequently, the distribution of $\bar{X}$ has mean

$$\mu_{\bar{X}} = 1(.0625) + 1.5(.1250) + \cdots + 3.5(.1250) + 4(.0625) = 2.5$$

which may be obtained directly from the symmetry of the distribution. The distribution of $\bar{X}$ has variance

$$\sigma^2_{\bar{X}} \;=\; (1 - 2.5)^2(.0625) + (1.5 - 2.5)^2(.1250) + \cdots$$

$$+ (3.5 - 2.5)^2(.1250) + (4 - 2.5)^2(.0625)$$

$$= .625.$$

Now the mean of the original distribution is also 2.5 and the variance is 1.25. Thus, Theorem 6.1 yields 2.5 and $1.25/2 = .625$ as the mean and the variance of the distribution of the sample mean of two observations. These agree exactly as they must.

6.10 The 25 means are

3.80, 4.25, 4.25, 5.05, 4.85,
4.20, 4.05, 4.15, 4.85, 4.20,
3.00, 5.20, 4.30, 4.45, 3.75,
5.40, 5.60, 5.70, 4.00, 5.10,
3.15, 4.50, 3.40, 5.00, 4.50.

The sample mean of these 25 means is 4.428.
The sample standard deviation of these 25 means is .714.
The population mean is:

$$0\frac{1}{10} + 1\frac{1}{10} + \cdots + 9\frac{1}{10} = 4.5.$$

The population variance is:

$$0^2\frac{1}{10} + 1^2\frac{1}{10} + \cdots + 9^2\frac{1}{10} - (4.5)^2 = 8.25,$$

so the population standard deviation is 2.87228. Theorem 6.1 says that the mean of the distribution of $\bar{X}$ is 4.5 and the variance is $8.25/20 = .4125$ Thus the theoretical standard deviation is .642. These compare well to the sample results.

6.11 The variance of the sample mean $\bar{X}$, based on a sample of size n, is σ^2/n. Thus the standard deviation, or standard error of the mean is $\sigma/\sqrt{n}$.

(a) The standard deviation for a sample of size 50 is $\sigma/\sqrt{50}$. The standard deviation for a sample is size 200 is $\sigma/\sqrt{200}$. That is, the ratio of standard errors is

$$\frac{\sigma/\sqrt{200}}{\sigma/\sqrt{50}} = \frac{\sqrt{50}}{\sqrt{200}} = \frac{1}{2},$$

so the standard error is halved.

(b) The ratio of standard errors is

$$\frac{\sigma/\sqrt{900}}{\sigma/\sqrt{400}} = \frac{\sqrt{400}}{\sqrt{900}} = \frac{2}{3},$$

so the standard error for a sample size 900 is 2/3rd's that for sample size 400.

(c) The standard error for a sample size 25 is

$$\frac{\sqrt{225}}{\sqrt{25}} = 3$$

times as large as that for sample size 225.

(d) The standard error for a sample size 40 is

$$\frac{\sqrt{640}}{\sqrt{40}} = 4$$

times as large as that for sample size 640.

6.12 The finite population correction factor is given by $(N - n)/(N - 1)$. Thus, when

(a) $n = 5$ and $N = 200$, the finite population correction factor is

$$\frac{200 - 5}{200 - 1} = .9799.$$

(b) $n = 10$ and $N = 400$, the finite population correction factor is

$$\frac{400 - 10}{400 - 1} = .9774.$$

(c) $n = 100$ and $N = 5,000$, the finite population correction factor is

$$\frac{5,000 - 100}{5,000 - 1} = .9802.$$

6.13 We need to find $P(|\bar{X} - \mu| < .6745 \cdot \sigma/\sqrt{n})$. Since the standard deviation of the mean is $\sigma/\sqrt{n}$, the standardized variable $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ is approximately a normal random variable for large n (central limit theorem). Thus , we need to find:

$$P(|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}| < .6745).$$

Now, interpolating in Table 3 gives

$$P(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < .6745) = .75.$$

Thus,

$$P(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq -.6745) = P(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \geq .6745) = .25$$

so

$$P(|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}| < .6745) = .75 - .25 = .50.$$

The probability that the mean of a random sample of size n , from a population with standard deviation σ, will differ from μ by less than $(.6745)(\sigma/\sqrt{n})$ is approximately .5 for sufficiently large n.

6.14 (a) Chebyshev's theorem for $\bar{X}$ (which has standard deviation $\sigma/\sqrt{n}$) states that:

$$P(|\bar{X} - \mu| < \frac{k\sigma}{\sqrt{n}}) \geq 1 - \frac{1}{k^2}.$$

In this case, $\sigma = 2.4$ and $n = 25$ so we take $k = (1.2)5/2.4 = 2.5$. Then

$$P(|\bar{X} - \mu| < 1.2) = P(|\bar{X} - \mu| < \frac{2.5\sigma}{\sqrt{n}}) \geq 1 - \frac{1}{(2.5)^2} = .84.$$

(b) Under the central limit theorem

$$P(|\bar{X} - \mu| < \frac{k\sigma}{\sqrt{n}}) \approx F(k) - F(-k),$$

where F is the standard normal distribution in Table 3. Thus,

$$\begin{aligned} P(|\bar{X} - \mu| < 1.2) &= P(\frac{|\bar{X} - \mu|}{\sigma/\sqrt{n}} < 2.5) \approx F(2.5) - F(-2.5) \\ &= .9938 - .0062) = .9876. \end{aligned}$$

6.15 Here $\sigma/\sqrt{n} = 16/10 = 1.6$ and

$$P(175 \leq \bar{X} \leq 178) = P(-1 \leq \bar{X} - 176 \leq 2) = P(-.625 \leq \frac{\bar{X} - 176}{16/10} \leq 1.25).$$

Since $n = 100$ is large , the central limit theorem yields the approximation

$$\begin{aligned} P(-.625 \leq \frac{\bar{X} - 176}{16/10} \leq 1.25) &\approx F(1.25) - F(-.625) \\ &= .8944 - .266) = .628. \end{aligned}$$

6.16 Since the population is normal, the distribution of $\bar{X}$ is exactly normal and we need to find

$$P(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{7.75 - 10}{1.5/\sqrt{4}}) = P(\frac{\bar{X} - 10}{1.5/2} < -3.00) = F(-3.00) = .0013.$$

Thus, if the process is deemed "out of control" by the decision rule (i.e. $\bar{X} < 7.75$ means that the process is termed "out of control"), then it is very unlikely that the process is actually in control. However, the process may actually be out of control with relatively high probability of not being detected.

6.17 We need to find

$$\begin{aligned} P(\sum_{i=1}^{36} X_i > 6,000) &= P(\bar{X} > 166.67) = P(\bar{X} - 163 > 3.67) \\ &= P(\frac{\bar{X} - 163}{18/6} > 1.222). \end{aligned}$$

Since $n = 36$ is relatively large, we use the central limit theorem to approximate this probability by

$$1 - F(1.222) = .111.$$

6.18 (a) If X has density $f(x)$, and $Y = X - \mu$, then

$$\mu_Y = \int_{-\infty}^{\infty} (x - \mu) f(x)\, dx = \int_{-\infty}^{\infty} x f(x)\, dx - \mu \int_{-\infty}^{\infty} f(x)\, dx.$$

But,

$$\int_{-\infty}^{\infty} x f(x)\, dx = \mu, \quad \text{and} \quad \int_{-\infty}^{\infty} f(x)\, dx = 1,$$

so,

$$\mu_Y = \mu - \mu = 0.$$

(b)

$$\begin{aligned} \sigma_Y^2 &= \text{expected value of } (Y - \mu_Y)^2 \\ &= E((X - \mu) - 0)^2 \\ &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx \;=\; \sigma^2. \end{aligned}$$

6.19 Let $\mu_X = E(X)$ be the expected value of X. First we will show that $E(X + Y) =$

$E(X) + E(Y)$. Let $f(x, y)$ be the joint density function of X and Y. Then, if X takes on discrete values x_i and Y takes on discrete values y_j,

$$\begin{aligned} E(X+Y) &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}(x_i + y_j)f(x_i, y_j) \\ &= \sum_{i=0}^{\infty} x_i \sum_{j=0}^{\infty} f(x_i, y_j) + \sum_{j=0}^{\infty} y_j \sum_{i=0}^{\infty} f(x_i, y_j). \end{aligned}$$

But, $\sum_{j=0}^{\infty} f(x_i, y_j) = f_1(x)$, the marginal density of X. Similarly for Y. Thus

$$E(X+Y) = \sum_{i=0}^{\infty} x_i f_1(x_i) + \sum_{j=0}^{\infty} y_j f_2(y_j) = E(X) + E(Y)$$

by definition. By induction

$$E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i).$$

Thus,

$$\begin{aligned} \mu_{\bar{X}} &= E(\frac{\sum_{i=1}^{n} X_i}{n}) \\ &= \frac{1}{n}E(\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i). \end{aligned}$$

But, each $E(X_i) = \mu$ so $\mu_{\bar{X}} = \mu$.

6.20 Since the sample is from a normal population

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{47.5 - 42.1}{8.4/5} = 3.214$$

is the value of a t random variable with 24 degrees of freedom. From Table 4,

$$P(\frac{\bar{X} - \mu}{S/\sqrt{n}} > 2.797) = .005.$$

Before sampling, the probability of getting a value of $t > 2.797$ is .005. The probability of being as large or larger than the observed $t = 3.214$ is even smaller. Consequently, the data tends to refute the claim that the population mean is 42.1 . It is likely that the true mean is greater then 42.1.

6.21 The mean of the data is $\bar{x} = 23$ and the sample standard deviation is 6.39. Thus, if the data is from a normal population with $\mu = 20$, the statistic

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{23 - 20}{6.39/\sqrt{6}} = 1.15$$

is the value of a t random variable with 5 degrees of freedom. The entry in Table 4 for $\alpha = 10$ and $\nu = 5$ is 1.476. Before the data are observed, we know that

$$P(\frac{\bar{X} - \mu}{S/\sqrt{n}} > 1.15) > .10.$$

Thus, the data does not give strong evidence against the ambulance service's claim.

6.22 If the sample is from a normal population with $\mu = .5000$

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{.5060 - .5000}{.0040/\sqrt{10}} = 4.74$$

is the value of a t random variable with 9 degrees of freedom. From Table 4,

$$P(\frac{\bar{X} - \mu}{S/\sqrt{n}} > 3.25) = .005.$$

Before observing the data, the probability of getting a value of $t > 3.25$ is .005. The probability of being as large or larger than the observed $t = 4.74$ is even smaller. Thus it is very unlikely that the process is under control. The data suggest that the true mean is greater then .5000 cm.

6.23 We need to find

$$P(S^2 > 39.74) \;=\; P(\frac{(n-1)S^2}{\sigma^2} > \frac{(n-1)39.74}{\sigma^2})$$

$$= P(\frac{(n-1)S^2}{\sigma^2} > \frac{(14)39.74}{21.32}) \quad = \quad P(\frac{14S^2}{\sigma^2} > 26.12).$$

Since the data are from a normal population, the statistic $(n-1)S^2/\sigma^2$ has a χ^2 distribution with $n-1$ degrees of freedom. From Table 5, with $\nu = 14$ degrees of freedom, we see that

$$P(\frac{14S^2}{\sigma^2} > 26.12) = .025.$$

Thus, the probability that the claim will be rejected even though $\sigma^2 = 21.3$ is .025 or 2.5 percent.

6.24 We need to find

$$P(3.14 < S^2 < 8.94) = P(\frac{9(3.14)^2}{42.5} < \frac{(n-1)S^2}{\sigma^2} < \frac{9(8.94)^2}{42.5}).$$

Since the data are from a normal population, the statistic $(n-1)S^2/\sigma^2$ has a χ^2 distribution with $\nu = n-1$ degrees of freedom. From Table 5, with $\nu = 9$ degrees of freedom, we see that

$$\begin{aligned} P(3.14 < S^2 < 8.94) &= P(\frac{(n-1)S^2}{\sigma^2} < 16.925) - P(\frac{(n-1)S^2}{\sigma^2} < 2.0879) \\ &= .95 - .01 \quad = \quad .94. \end{aligned}$$

6.25 We need to find

$$1 - P(\frac{1}{7} \le \frac{S_1^2}{S_2^2} \le 7).$$

Since the samples are independent and from normal populations, the statistic S_1^2/S_2^2 has an F distribution with $n_1 - 1 = n_2 - 1 = 7$ degrees of freedom for both the numerator and the denominator. From Table 6(b), we see that

$$P(\frac{S_1^2}{S_2^2} > 6.99) = .01.$$

Using the relation

$$F_{1-\alpha}(\nu_1,\nu_2) = \frac{1}{F_\alpha(\nu_2,\nu_1)},$$

we know that

$$P(\frac{S_1^2}{S_2^2} < \frac{1}{6.99}) = .01.$$

Thus, approximately

$$P(\frac{S_1^2}{S_2^2} < \frac{1}{7} \; or \; \frac{S_1^2}{S_2^2} > 7)$$

$$= P(\frac{S_1^2}{S_2^2} < \frac{1}{7}) + P(\frac{S_1^2}{S_2^2} > 7) = .01 + .01 + = .02.$$

6.26 (a)

$$F_{.95}(12,15) = \frac{1}{F_{.05}(15,12)} = \frac{1}{2.62} = .382.$$

(b)

$$F_{.99}(6,20) = \frac{1}{F_{.01}(20,6)} = \frac{1}{7.40} = .135.$$

6.27 We need to find

$$P(S^2 > 180) = P(\frac{(n-1)S^2}{\sigma^2} > \frac{4 \cdot 180}{144}) = P(\frac{(n-1)S^2}{\sigma^2} > 5).$$

Since the sample is from a normal population, the statistic $((n-1)S^2)/(\sigma^2)$ has a χ^2 distribution with $n-1=4$ degrees of freedom. The density of the χ^2 distribution with 4 degrees of freedom is

$$f(x) = \frac{1}{4}xe^{-x/2}, \qquad x > 0.$$

Thus,

$$P(S^2 > 180) = P(\frac{(n-1)S^2}{\sigma^2} > 5) \; = \; \int_5^\infty \frac{1}{4}xe^{-x/2}\,dx.$$

Integrating by parts with $u = x$ and $dv = e^{-x/2}$ gives

$$P(S^2 > 180) = -e^{-x/2}(\frac{1}{2}x + 1) \mid_5^{\infty} = e^{-5/2}(\frac{5}{2} + 1) = e^{-5/2}\frac{7}{2} = .2873.$$

6.28 The value for $t_{.05}$ for $\nu = 1$ in Table 4 is 6.314. Thus we need to show that

$$P(t > t_{.05}) = P(t > 6.314) = \int_{6.314}^{\infty} \frac{1}{\pi}\frac{1}{1+t^2}dt = .05.$$

But,

$$\int_{6.314}^{\infty} \frac{1}{\pi}\frac{1}{1+t^2}dt = \frac{1}{\pi}\arctan(t) \mid_{6.314}^{\infty} = \frac{1}{\pi}(\frac{\pi}{2} - 1.4137) = .05$$

as required.

6.29 The probability that the ratio of the larger to the smaller sample variance exceeds 3 is

$$1 - P(\frac{1}{3} \leq \frac{S_1^2}{S_2^2} \leq 3) = 1 - \int_{1/3}^{3} \frac{6x}{(1+x)^4}\, dx.$$

Let $u = x + 1$ or $x = u - 1$. Then

$$\begin{aligned} P(\frac{1}{3} \leq \frac{S_1^2}{S_2^2} \leq 3) &= \int_{4/3}^{4} \frac{6(u-1)}{u^4}\, du = \frac{-3}{u^2} \mid_{4/3}^{4} + \frac{2}{u^3} \mid_{4/3}^{4} \\ &= \frac{24}{16} - \frac{52}{64} = .6875. \end{aligned}$$

Thus,

$$1 - P(\frac{1}{3} \leq \frac{S_1^2}{S_2^2} \leq 3) = 1 - .6875 = .3125.$$

6.30 (a) Number the states from 1 to 50 according to their alphabetical order. Using the last two columns of the second page of Table 7, starting at row 11, to select 10 numbers between 1 and 50 by discarding numbers outside of this range as well as any previously drawn number gives

17, 2, 12, 1, 44, 18, 33, 39, 40, 41.

These correspond to Kentucky, Alaska, Idaho, Alabama, Utah, Louisiana, North Carolina , Rhode Island, South Carolina, and South Dakota.

(b) No. High population states like California would have many participants but could send only two. A student from a small state would have a better chance than a student in California.

6.31 (a) The number of samples (given that order does not matter) is

$$\binom{8}{2} = \frac{8 \cdot 7}{2 \cdot 1} = 28.$$

(b) The number of samples (given that order does not matter) is

$$\binom{20}{2} = \frac{20 \cdot 19}{2 \cdot 1} = 190.$$

6.32 For a random sample, each outcome must have equal probability. The probabilities are :

(a) $\frac{1}{28} = .0357.$

(b) $\frac{1}{190} = .00526.$

6.33 The finite population correction factor is $(N-n)/(N-1)$. Thus

(a)

$$\frac{N-n}{N-1} = \frac{8-2}{8-1} = .857.$$

(b)

$$\frac{N-n}{N-1} = \frac{20-2}{20-1} = .947.$$

6.34 (a) Here $\sigma/\sqrt{n} = 9/6 = 1.5$ and we will bound the one-tailed probability

$$P(\bar{X} > 66.75) = P(\bar{X} - 63 > 66.75 - 63).$$

Using Chebyshev's theorem for the two-tailed probability

$$\begin{aligned} P(|\bar{X} - 63| > 66.75 - 63) &= P(|\bar{X} - 63| > \frac{66.75 - 63}{9/6} \cdot 9/6) \\ &\leq (\frac{9/6}{66.75 - 63})^2 = .16. \end{aligned}$$

Thus, $P(\bar{X} > 66.75) \leq .16$.

(b) We need to find

$$P(\frac{\bar{X} - 63}{\sigma/\sqrt{n}} > \frac{66.75 - 63}{9/6}) = P(\frac{\bar{X} - 63}{\sigma/\sqrt{n}} > 2.5)$$

which by the central limit theorem is

$$P(Z > 2.5) = 1 - F(2.5) = 1 - .9938 = .0062.$$

6.35 Let $\bar{X}$ be the mean number of pieces of mail in a random sample of size 25.

(a) By Chebyshev's theorem, set

$$P(\bar{X} < 40 \text{ or } \bar{X} > 48) = P(|\bar{X} - 44| > 4) \leq \frac{1}{k^2}$$

where

$$4 = k\sigma_{\bar{X}} = k\frac{8}{\sqrt{25}} \quad \text{or} \quad \frac{1}{k^2} = .16$$

Thus,

$$P(\bar{X} < 40 \text{ or } \bar{X} > 48) \leq .16$$

(b) By the central limit theorem,

$$\begin{aligned} P(\bar{X} < 40 \text{ or } \bar{X} > 48) &= 1 - P(40 \le \bar{X} \le 48) \\ &= 1 - P(\frac{40-44}{8/\sqrt{25}} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le \frac{48-44}{8/\sqrt{25}}) \\ &\approx 1 - P(-2.5 \le Z \le 2.5) = 2P(Z > 2.5) \\ &= 2(1 - .9938) = .0124 \end{aligned}$$

6.36 Since the population is normal, the distribution of $\bar{X}$ is exactly normal and we need to find

$$\begin{aligned} P(|\bar{X} - \mu| \le .02) &= P(|\frac{\bar{X}-\mu}{.04/5}| \le \frac{.02}{.04/5}) = F(2.5) - F(-2.5) \\ &= .9938 - .0062) = .9876. \end{aligned}$$

6.37 By the central limit theorem,

$$\begin{aligned} P(|\bar{X} - \mu| > .06) &= P(\frac{|\bar{X}-\mu|}{\sigma/\sqrt{n}} > \frac{.06}{.16/\sqrt{36}}) \\ &\approx P(|Z| > 2.25) = 2(1 - .9878) = .0244 \end{aligned}$$

6.38 We need to find

$$P(\frac{S_1^2}{S_2^2} > 4.0).$$

Since the samples are independent and from normal populations, the statistic S_1^2/S_2^2 has an F distribution with $n_1 - 1 = 8$ and $n_2 - 1 = 15$ degrees of freedom for the numerator and the denominator, respectively . From Table 6(b), we see that

$$P(\frac{S_1^2}{S_2^2} > 4.0) = .01,$$

so the probability is .01 or 1 percent.

6.39 By Theorem 6.5, $F = S_2^2/S_1^2$ has F distribution with 7 and 25 degrees of freedom.

Hence , from Table 6(a), we have

$$P(\frac{S_2^2}{S_1^2} > 2.4) = P(F > 2.4) = .05$$

6.40 The variance of the sample mean $\bar{X}$, based on a sample of size n, is σ^2/n. Thus the standard deviation, or standard error, of the mean is $\sigma/\sqrt{n}$.

(a) The standard deviation for a sample of size 100 is $\sigma/\sqrt{100}$. The standard deviation for a sample is size 200 is $\sigma/\sqrt{200}$. That is, the ratio of standard errors is

$$\frac{\sigma/\sqrt{200}}{\sigma/\sqrt{100}} = \frac{\sqrt{100}}{\sqrt{200}} = \frac{1}{\sqrt{2}} = .707,$$

so the standard error for sample size 200 is .707 times the standard error for sample size 100.

(b) The ratio of standard errors is

$$\frac{\sigma/\sqrt{300}}{\sigma/\sqrt{200}} = \frac{\sqrt{200}}{\sqrt{300}} = \sqrt{\frac{2}{3}} = .816,$$

so the standard error for a sample size 300 is .816 times that for sample size 200.

(c) The ratio of standard errors is

$$\frac{\sigma/\sqrt{90}}{\sigma/\sqrt{360}} = 2,$$

so the standard error doubles.

6.41 Under this sampling scheme, the observations will not satisfy the independence assumption required for a random sample. The longest lines are likely to occur together when the same cars must wait several light changes during the rush hour.

6.42 (a) The production process may not be stabilized when the first shaft is made each morning. Alternatively, the first shaft could be given extra attention and may not be representative.

(b) Buses and large trucks would take longer to pass the fixed point and thus would be more likely to be included in the sample than a small car.

6.43 (a) No, we would expect the weight of the bags to vary because the weights of the individual apples vary. With X_i the weight of the $i-th$ apple placed in the bag, $\sum_{i=1}^{3} 0X_i$ is the total weight of the bag. Then,

$$P(\sum_{i=1}^{3} 0X_i \ \leq y) \ = \ P(\overline{X} \leq y/30)$$

so the probability concerning total weight can be expressed as a probability concerning $\overline{X}$.

(b) Further,

$$P(\overline{X} \leq y/30) = \frac{y/30 - \mu}{\sigma/\sqrt{n}} \ )$$

By the central limit theorem $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ is nearly standard normal when $n =$ 30. Thus, for any specified value y, we can approximate the probability $P(\sum_{i=1}^{30} X_i \ \leq y)$ by the standard normal probability $P(\ Z \leq \frac{y/30 - \mu}{\sigma/\sqrt{n}} \ )$

(c) With $\mu = .2$ and $\sigma = .03$ pound

$$P(\ \sum_{i=1}^{30} X_i \ \geq 6.2) \ = \ P(\ Z \geq \frac{6.2/30 - .2}{.03/\sqrt{30}} \ ) = 1 - F(1.217) = .1118$$