Homework Assignment 2

(Due in class on Tuesday, 01/28/2014)

STA 371G, Statistics and Modeling, Spring 2014

Problem 1

Suppose $Z \sim \mathcal{N}(0, 1)$, i.e., the random variable Z follows a standard normal distribution.

(a) Find $P(Z < -1.25)$.

$P(Z < -1.25) = 0.106$
In Excel: =NORMSDIST(-1.25)
In R: pnorm(-1.25)

(b) Find $P(Z > 1.50)$.

$P(Z > 1.50) = 1 - P(Z \leq 1.50) = 0.067$

(c) Find $P(-0.10 < Z < 1.40)$.

$P(-0.10 < Z < 1.40) = P(Z < 1.40) - P(Z \leq -0.10) = 0.459$

Problem 2

Suppose $X \sim \mathcal{N}(5, 10)$, i.e., X is normal distributed with mean 5 and variance 10. Compute:

(a) $P(X > 5)$

$$P(X > 5) = P\left(Z > \frac{5-5}{\sqrt{10}}\right) = P(Z > 0) = 0.50 \quad \text{(or 50\%)}$$

You can do this without using Excel or R, looking at the normal table or using your calculator... just remember that in the normal distribution, 50% of the probability is above the mean.

(b) $P(X > 5 + 2 \times \sqrt{10})$

$$P(X > 5 + 2 \times \sqrt{10}) \approx 0.025 \quad \text{(or 2.5\%)}$$

Rule of thumb: In the normal distribution, about 95% of the probability is between -2 and 2 standard deviations.

You can also use Excel: = 1-NORMDIST(5+2*sqrt(10),5,sqrt(10),TRUE)
or R: 1-pnorm(5+2*sqrt(10),5,sqrt(10))

(c) $P(X = 8)$

$P(X = 8) = 0$

(d) Express $P(-2 \le X \le 6)$ in terms of Z, the standard normal random variable.

$$P(-2 \le X \le 6) = P\left(\frac{-2-5}{\sqrt{10}} \le Z \le \frac{6-5}{\sqrt{10}}\right) = 0.61$$

Problem 3

Suppose that Z follows a standard normal distribution. The probability is 10% that Z is greater than what number?

We are trying to find $P(Z >?) = 0.10$, i.e., $P(Z \le ?) = 1 - 0.10 = 0.90$.
In Excel: =NORMSINV(0.9)
In R: qnorm(0.90)
The answer is found to be 1.282.

Problem 4

A company can purchase raw material from either Supplier A or Supplier B and is concerned about the amounts of impurity the material contains. A review of the records for each supplier indicates that the percentage impurity levels in consignments of the raw material follow normal distributions with the means and standard deviations given in the table below. The company is particularly anxious that the impurity level in a consignment not exceed 5% and want to purchase from the supplier more likely to meet that specification. Which supplier should be chosen?

	Mean	Standard Deviation
Supplier A	4.4	0.4
Supplier B	4.2	0.6

Let X_A represent the percentage of impurity level in a randomly chosen consignment of raw material from Supplier A. Therefore, $X_A \sim N\left(4.4, 0.4^2\right)$. Similarly, X_B represents the percentage of impurity level in a randomly chosen consignment of raw material from Supplier B, and, $X_B \sim \left(4.2, 0.6^2\right)$.

We need to compute $P(X_A > 5)$ and $P(X_B > 5)$.

$$P(X_A > 5) = P\left(Z > \frac{5-4.4}{0.4}\right) = P(Z > 1.50)$$

and

$$P(X_B > 5) = P\left(Z > \frac{5-4.2}{0.6}\right) = P(Z > 1.33)$$

Since $P(Z > 1.50) < P(Z > 1.33)$, we conclude that Supplier A is better and should be chosen.

Problem 5

The tread life of a particular brand of tires has a normal distribution with mean 40,000 miles and standard deviation 5,000 miles. What proportion of these tires will have tread lives between 35,000 and 45,000 miles? What proportion of these tires will have tread lives between 30,000 and 50,000 miles?

Since the tread life $X \sim \mathcal{N}(40000, 5000^2)$ (X is normal distributed with mean 40000 and variance 5000^2).

$$P(35000 \leq X \leq 45000) = P\left(\frac{35000-40000}{5000} \leq Z \leq \frac{45000-40000}{5000}\right) = P(-1 \leq Z \leq 1) = 0.68$$

$$P(30000 \leq X \leq 50000) = P\left(\frac{30000-40000}{5000} \leq Z \leq \frac{50000-40000}{5000}\right) = P(-2 \leq Z \leq 2) = 0.95$$

You can do this without using software. Rule of thumb: In the normal distribution, about 68% of the probability is between -1 and 1 standard deviations, and about 95% of the probability is between -2 and 2 standard deviations

Problem 6

According to the official Federal Election Commission report for the presidential election in 2012, out of a total of 7,993,851 votes in Texas, President Barack Obama received 3,308,124 votes.

(a) If you randomly survey 1000 Texas residents who had voted in the 2012 presidential election, can you predict the distribution of the number of votes for President Obama among these 1000 Texas voters? Will you be surprised to find out that more than 500 of them voted for President Obama? (Hint: using the normal approximation to the binomial distribution.)

The proportion of votes for President Obama in Texas in the 2012 election is

$$p = \frac{3,308,124}{7,993,851} = 0.414 \text{ (or 41.4\%)}.$$

As the number of voters $7,993,851$ is much larger than the sample size $n = 1000$, we can consider the number of votes for President Obama in a random sample of $n = 1000$ to be binomial distributed as

$$X \sim \text{Binomial}(n = 1000, p = 0.414).$$

Since n is large and p is close to neither 0 nor 1, we may safely approximate the binomial distribution with a normal distribution with mean np and variance $np(1-p)$, which can be expressed as

$$X \sim \mathcal{N}(np, np(1-p)) = \mathcal{N}(413.8, 15.6^2).$$

Thus we are 95% confident that the number of votes for President Obama from a random sample of $n = 1000$ Texas voters is between

$$[382, 445]$$

Note that $X \geq 500$ is outside $[382, 445]$. Further calculating $P(X \geq 500)$ using

$$1 - pnorm(500, 413.8, 15.6)$$

in R, or

$$1 - NORMDIST(500, 413.8, 15.6, TRUE)$$

in Excel, we find that $P(X \geq 500)$ is close to zero, thus we would be very surprised to find out $X \geq 500$.

(b) According to http://www.politico.com/2012-election/results/president/texas/, In Dallas County, TX, President Obama received 57.1% of the votes in the 2012 presidential election. If you randomly survey 100 residents of Dallas County who had voted in the 2012 presidential election, can you predict the distribution of the number of votes for President Obama among them? Will you be surprised to find out no more than 50 votes for President Obama?

Similar to the analysis for Problem 6.(b), the number of votes X can be considered distributed as

$$X \sim \text{Binomial}(n = 100, p = 0.571)$$

Approximately, we have

$$X \sim \mathcal{N}(100 * 0.571, 100 * 0.571 * (1 - 0.571)) = \mathcal{N}(57.1, 4.95^2).$$

Thus we are 95% confident that the number of votes for President Obama from a random sample of $n = 100$ Dallas County voters is between

$$[47, 67]$$

Note that $X = 50$ is between [47, 67]. Further calculating $P(X \leq 50)$ using

$$pnorm(50, 57.1, 4.95)$$

in R, or

$$1 - NORMDIST(50, 57.1, 4.95, TRUE)$$

in Excel, we find that

$$P(X \leq 50) = 7.6\%.$$

As $P(X \leq 50)$ happen with a probability that is not small, we are usually not surprised to find out that X is no more than 50.