INTRODUCTION TO STATISTICS FALL 2010

Exercise 10

1. An expert witness in a paternity suit testifies that the length (in days) of pregnancy (that is, the time from impregnation to the delivery of the child) is approximately normally distributed with parameter $\mu = 270$ and $\sigma^2 = 100$. The defendant in the suit is able to prove that he was out of the country during a period that began 290 days before the birth of the child and ended 240 days before the birth. If the defendant was, in fact, the father of the child, what is the probability that the mother could have had a very long or very short pregnancy indicated by the testimony?

   Want P(X < 240) and P(X > 290).

   $$P(X < 240) = P(Z < \frac{240-270}{10}) = P(Z < -3) = P(Z > 3)$$
   $$= 1 - P(Z < 3) = 1 - 0.9987 = 0.0013$$

   $$P(X > 290) = P(X > 290) = P(Z > \frac{290270}{10} = P(Z > 2)$$
   $$= 1 - P(Z < 2) = 1 - 0.9772 = 0.0228$$

   Combined $P = 0.0013 + 0.0228 = \mathbf{0.0241} < 0.05$, very unlikely assuming normal distribution is true.

2. Early in the morning, a group of m people decides to use the elevator in an otherwise deserted building of 21 floors. Each of these persons chooses his or her floor independently of the others, and - from our point of view - completely random, so that each person selects a floor with probability 1/21. Let S_m be the number of times the elevator stops. In order to study S_m, we introduce for $i = 1, 2, \ldots, 21$ random variables R_i, given by

   $$R_i = \begin{cases} 1 & \text{if the elevator stops at the } i\text{th floor,} \\ 0 & \text{if the elevator does not stop at the } i\text{th floor.} \end{cases}$$

   (a) Each R_i has $Ber(p)$ distribution. Calculate the parameter p. A discrete random variable X has Bernoulli distribution with parameter p, where $0 \leq p \leq 1$, if its probability mass function is given by $P(X = 1) = p$ and $P(X = 0) = 1p$. We denote it by $Ber(p)$.

   We are ONLY dealing with floors!
   N = 21. So we define $P(R_i = 0)$ = P(elevator did not stop at floor i |m people in elevator) = P(All m people in elevator did NOT select floor i) = $(1 - \frac{1}{21})^m = (\frac{20}{21})^m$
   $\mathbf{P(R_i = 1) = p = 1 - P(R_i = 0) = 1 - (\frac{20}{21})^m}$

   (b) From the way we defined S_m, Can we conclude that S_m has $Bin(21, p)$ distribution, with p as in part a?. Why or why not?

   Your answer here.

(c) $P(S_2 = 1) = \dfrac{1}{21} = 1 - P(S_2 = 2)$
When there are m=2 people, elevator has max. 2 stops (and min. 1)

$P(S_2 = 2) = \dfrac{20}{21}$
Since $1 = P(S_2 = 1) + P(S_2 = 2)$.

$P(S_3 = 1) = \dfrac{21}{21}\left(\dfrac{1}{21}\right)^2 = \dfrac{1}{441}$
First person has 21/21 floors to choose from. If elevator makes only one stop for 3 people, than 2nd and 3rd person got off the same floor (that is, they only had 1 available choice of the 21)

$P(S_3 = 2) = \dfrac{21}{21}\left(3 \times \dfrac{1}{21}\dfrac{20}{21}\right) = \dfrac{60}{441}$
3 possibilities: 1st and 2nd person get off same stop, 1st and 3rd get off same stop, and 2nd + 3rd get off same stop. In last possibility, 2nd has 20/21 choices different from 1st, and 3rd only has 1 choice which is the same as the 2nd.

$P(S_3 = 3) = \dfrac{21}{21}\dfrac{20}{21}\dfrac{19}{21} = \dfrac{380}{441}$
Different floors selected by each person.

3. Based on past experience, a bank belives that 7% of the people who receive loans will not make payments on time. The bank has recently approved 200 loans.

   (a) What is the mean and standard deviation of the proportion of clients in this group who may not make timely payments?

      Key words are *proportion* and *in this group*. We are looking for the proportion of sample mean and sample standard deviation. By the formulas on p.316 in the textbook:
      E(proportion of recent loans not paid on time) $= E(\bar{X}) = p = \mathbf{0.07}$
      $\text{SD}(\bar{X}) = \sqrt{\dfrac{p(1-p)}{n}} = \sqrt{\dfrac{0.07(0.93)}{200}} \approx \mathbf{0.018}$

   (b) What assumptions underlie the model? Are the conditions met? Explain.

      If we define X = number of ALL loans with bank (population) that will not be paid on time, then X is approximately a binomial distribution with parameters n and p provided n is lage enough. That is, when both np and $n(1-p) > 5$.
      $np = 200 * 0.07 = 14 > 5$
      $n(1-p) = 200 * 0.93 = 86 > 5$

   (c) What is the probability that over 10% of these clients will not make timely payments?

We want $P(X > 0.10) = P(Z > \frac{0.10 - 0.07}{0.018}) = P(Z > 1.667) = 1 - P(Z < 1.667) = \mathbf{0.0485}$