STAT 6200 Homework #3 - Solution

9. X = # of days out of $n=7$ on which the concentration of CO surpasses a specified level.
Several properties of a binomial experiment are likely to be violated here. ~~Hence~~ The seven days constitute repeated trials, but they are not conducted under identical circumstances. E.g., the weather is likely to differ from day to day. ~~This may~~ In addition the # of cars on the roads on Saturday + Sunday may be quite different than on weekdays. As a consequence of this, the probability of a "success" (a day where CO concentration exceeds the threshold) may differ from day to day, or at least between weekdays + weekends. Finally, and perhaps most clearly, the outcomes on consecutive days are probably not independent. That is, if the CO level is high on Monday, it will be more likely to be high on Tuesday because CO lingers in the air.

11. a) $n!$ gives the # of ways that n items can be ordered. So the answer here is $10! = 3{,}628{,}800$

b) $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ gives the # of ways that x items can be chosen, without regard to order, from n items.

So, the answer here is $\binom{10}{4} = \frac{10!}{4!\,6!} = \frac{10\cdot9\cdot8\cdot7\cdot\cancel{6}\cdot\cancel{5}\cdot\cancel{4}\cdot\cancel{3}\cdot\cancel{2}\cdot\cancel{1}}{4\cdot3\cdot2\cdot1\cdot\cancel{6}\cdot\cancel{5}\cdot\cancel{4}\cdot\cancel{3}\cdot\cancel{2}\cdot\cancel{1}}$

$= 210$

c.) If the 10 people are chosen at random then this is a binomial experiment where X = # that are left-handed out of 10 is distributed as

$$X \sim Bin\,(n=10,\ p=.098)$$

$$P(X = \overset{x}{\downarrow}3) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{10}{3} (.098)^3 (1-.098)^{10-3}$$

$$= \underbrace{\frac{10!}{3!\,7!}}_{=\frac{10\cdot 9\cdot 8}{3\cdot 2\cdot 1} = 120} (.098)^3 (.902)^7 = .0549$$

d.) $P(X \geq 6) = P(X=6) + P(X=7) + P(X=8) + P(X=9) + P(X=10)$

$$= \binom{10}{6}(.098)^6(.902)^4 + \binom{10}{7}(.098)^7(.902)^3 + \binom{10}{8}(.098)^8(.902)^2$$

$$+ \binom{10}{9}(.098)^9(.902)^1 + \binom{10}{10}(.098)^{10}(.902)^0$$

$$= \cancel{.00012}\ .0001231 + .0000076 + .0000003 + .000000008$$
$$+ 8.2 \times 10^{-11} = .000131$$

Could also do this by using

$$P(X \geq 6) = 1 - \underbrace{P(X \leq 5)}_{=.999869 \text{ from computer}} = 1 - .999869 = .000131$$

e) $P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) = .3565 + .3873 + .1894$
$= .9332$

12. a) Let X = # of individuals w/ sedentary lifestyle among 12 randomly selected from the US population

$$X \sim Bin(\underset{=n}{12}, \underset{=p}{.58})$$

$E(X) = np = 12(.58) = 6.96$, $Var(X) = np(1-p) = 12(.58)(.42) = 2.9232$

$SD(X) = \sqrt{Var(X)} = \sqrt{2.9232} = 1.71$

b.) $P(X \geq 10) = P(X=10) + P(X=11) + P(X=12)$

$= \binom{12}{10}(.58)^{10}(.42)^{2} + \binom{12}{11}(.58)^{11}(.42)^{1} + \binom{12}{12}(.58)^{12}(.42)^{0}$

$= .05016 + .01259 + .00145 = .0642$

14. X = # cases in a month, $X \sim Poisson(\lambda)$ where $\lambda = 4.5$

a) $P(X = \underset{x}{1}) = \dfrac{e^{-\lambda}\lambda^{x}}{x!} = \dfrac{e^{-4.5}\, 4.5^{1}}{1!} = .0500$

b) $P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) = \dfrac{e^{-4.5}\, 4.5^{0}}{0!} + \dfrac{e^{-4.5}\, 4.5^{1}}{1!}$

$+ \dfrac{e^{-4.5}\, 4.5^{2}}{2!} = .0111 + .0500 + .1125 = .1736$

14 c) $P(X \geq 4) = 1 - P(X \leq 3) = 1 - [P(X=3) + \underbrace{P(X \leq 2)}_{\text{computed in part (b)}}]$

$$= 1 - \frac{e^{-4.5}\, 4.5^3}{3!} - .1736 = 1 - .1687 - .1736$$

$$= .6577$$

d.) $E(X) = \lambda = 4.5$, $SD(X) = \sqrt{var(X)} = \sqrt{\lambda} = \sqrt{4.5}$
$= 2.12$

16. We assume that $X \sim Bin(n, p)$ where $n = 2000$, $p = .0085$

a) $E(X) = np = 2000(.0085) = 17$

b.) This can be computed exactly from the binomial distribution with a computer program that gives cumulative probabilities. E.g., from Minitab we get

$P(X \leq 5) = .0006499.$

This value can also be approximated very well using the ~~normal~~ Poisson approximation to the binomial.

~~Here X has mean 17 and variance = np(1-p) = 2000(.0085)(1-.0085) = 16.8555~~
~~Therefore, the N(17, 16.8555) distribution is a can be good appr used to approximate the Bin(2000,.0085).~~

16 b.) (Continued) Let $Y \sim \text{Poisson}(\lambda)$ where $\lambda = np = 17$.

Then $P(X \le 5) \approx P(Y \le 5) = P(Y=0) + P(Y=1)$

$+ \cdots + P(Y=5) = \dfrac{e^{-17}\, 17^0}{0!} + \dfrac{e^{-17}\, 17^1}{1!} + \cdots + \dfrac{e^{-17}\, 17^5}{5!}$

$= \underbrace{.0000 + .0000 + .0000 + .0000 + .0001 + .0005}_{\text{from table in back of book}} = .0006$

close to true value of .0006499

c) $P(15 \le X \le 20) = P(X \le 20) - P(X < 15) = P(X \le 20) - P(X \le 14)$

$= .8064 - .2798 = .5266$

The Poisson(17) approximation gives

$P(15 \le X \le 20) \approx P(15 \le Y \le 20)$ for $Y \sim \text{Poisson}(17)$

$= P(Y \le 20) - P(Y \le 14) = .805481 - .280833$

$= .5246$ which is pretty close.

17. $Z \sim N(0,1)$

a) $P(Z \geq 2.60) = .005$ (back of book)

b.) $P(Z < 1.35) = 1 - P(Z \geq 1.35) = 1 - .089 = .911$

c) $P(-1.70 < Z < 3.10) = P(Z < 3.10) - P(Z < -1.70)$
$= [1 - P(Z > 3.10)] - P(Z > 1.70)$
$= 1 - .001 - .045 = .954$

d.) To get $Z_{.85}$ we look up .15 in the body of the table. .15 doesn't appear there, but .149 gives $Z_{.851} = 1.04$ and .152 gives $Z_{.848} = 1.03$, so $Z_{.85} \approx 1.03\cancel{6}$

e) $Z_{.20} = -Z_{.80}$ Looking up .20 in the table gives $Z_{.80} = .84$, so $Z_{.20} = -.84$

19. Let X = weight of US males (adults, presumably)
$X \sim N(172.2, (29.8)^2)$

a) $P(X < 130) = P\left(\frac{X-\mu}{\sigma} < \frac{130-\mu}{\sigma}\right) = P\left(Z < \frac{130-172.2}{29.8}\right)$
$= P(Z < -1.416) = P(Z > 1.416) \approx P(Z > 1.42) = .078$

back of our book

b.) $P(X > 210) = P\left(\frac{X-\mu}{\sigma} > \frac{210-\mu}{\sigma}\right) = P\left(Z > \frac{210 - 172.2}{29.8}\right)$

$= P(Z > 1.268) \approx P(Z > 1.27) = .102$

c) $P(X < 130 \cup X > 210) = P(X<130) + P(X>210) - \underbrace{P(X<130 \cap X>210)}_{=0}$

$= .078 + .102 = .180$

Let Y = # of males out of 5 that have weights outside of range 130 - 210 lbs

Then $Y \sim Bin(5, .180)$

$P(Y \geq 1) = 1 - P(Y=0) = 1 - \binom{5}{0}(.180)^0(1-.180)^{5-0}$

$= 1 - (.82)^5 = .629$

20 Let X = cholesterol for diseased

Y = " " not diseased

Then

$X \sim N(\mu_d = 244, \sigma_d^2 = (51)^2)$

$Y \sim N(\mu_{nd} = 219, \sigma_{nd}^2 = (41)^2)$

a) Given that we are testing men that will develop heart disease, the probability of concluding that the patient will develop heart disease is

$$P(X \geq 260) = P\left(\frac{X-\mu_d}{\sigma_d} \geq \frac{260-\mu_d}{\sigma_d}\right) = P\left(Z \geq \frac{260-244}{51}\right)$$

$$= P(Z \geq .31) = .378$$

b.) $$P(Y \geq 260) = P\left(\frac{Y-\mu_{nd}}{\sigma_{nd}} \geq \frac{260-\mu_{nd}}{\sigma_{nd}}\right) = P\left(Z \geq \frac{260-219}{41}\right)$$

$$= P(Z \geq 1) = .159$$

c) $$P(X < 260) = 1 - P(X \geq 260) = 1 - .378 = .622$$

d.) Prob of false positive $= P(\text{cholesterol} \geq \text{cutoff} \mid \text{not diseased})$
$= P(Y \geq \text{cutoff})$

Prob of false negative $= P(\text{cholesterol} < \text{cutoff} \mid \text{diseased})$
$= P(X < \text{cutoff})$

$P(Y \geq \text{cutoff}) = .159$ for cutoff $= 260$

$P(Y \geq 250) = P\left(Z \geq \frac{250-219}{41}\right) = P(Z \geq .756) = .224$ gets higher

$P(X < \text{cutoff}) = .622$ if cutoff $= 260$

$P(X < 250) = 1 - P(X \geq 250) = 1 - P\left(Z \geq \frac{250-244}{51}\right)$

$= 1 - P(Z \geq .12) = 1 - .452 = .548$ gets lower

20 e) initial cholesterol does not appear to be very useful because if we set the threshhold at 260 then we have a high probability of false negatives (.622). If we lower the threshhold, then we can reduce the false negative probability somewhat, but we increase the probability of a false positive to an unacceptably high level (.224 when the cutoff is 250). The ROC curve for initial cholesterol would not look good; it would ~~[illegible]~~ lie too close to the y=x diagonal line.