Nate Warner | STAT 300 | Summer 2024

PSET 8 - Due: Sunday, July 21

1. Let Z be the standard normal random variable. Use the table of Standard Normal Curve Areas to obtain each of the following probabilities.

   (a) $P(Z < -1.25)$
   (b) $P(Z > 2.48)$
   (c) $P(-2.71 < Z < 0.58)$
   (d) $P(|Z| \leqslant 2.50)$

a.) By symmetry, we have

$$P(Z < -1.25) = \Phi(-1.25) = 1 - \Phi(1.25) = 1 - 0.8944 = 0.1056.$$

b.)

$$P(Z > 2.48) = 1 - \Phi(2.48) = 1 - 0.9934 = 0.0066.$$

c.)

$$\begin{aligned} P(-2.71 < Z < 0.58) &= \Phi(0.58) - \Phi(-2.71) \\ &= \Phi(0.58) - (1 - \Phi(2.71)) \\ &= \Phi(0.58) - 1 + \Phi(2.71) \\ &= 0.719 - 1 + 0.9966 = 0.7156. \end{aligned}$$

d.)

$$\begin{aligned} P(|Z| \leqslant 2.5) &= P(-2.5 \leqslant Z \leqslant 2.5) \\ &= \Phi(2.5) - \Phi(-2.5) \\ &= \Phi(2.5) - (1 - \Phi(2.5)) \\ &= 0.9938 - (1 - 0.9938) \\ &= 0.9876. \end{aligned}$$

2. In each case, find the value of the constant c that makes the probability statement correct.

   (a) $P(Z \leqslant c) = 0.80$ (Note – the value c could also be described as the 80$^{\text{th}}$ percentile.)

   (b) $P(Z > c) = 0.025$

   (c) $P(0 < Z < c) = 0.291$

   (d) $P(-c < Z < c) = 0.668$

a.) By table A.3

$$\begin{aligned} P(Z \leqslant c) &= 0.8 \\ \implies c &= 0.84. \end{aligned}$$

b.)

$$\begin{aligned} P(Z > c) &= 0.025 \\ \implies 1 - P(Z < c) &= 0.025 \\ \implies P(Z < c) &= 0.975 \\ \implies c &= 1.96. \end{aligned}$$

c.)

$$\begin{aligned} P(0 < Z < c) &= 0.291 \\ \implies P(Z < c) - P(Z < 0) &= 0.291 \\ \implies \Phi(c) - \Phi(0) &= 0.291 \\ \implies \Phi(c) - 0.5 &= 0.291 \\ \implies \Phi(c) &= 0.791 \\ \implies c &= 0.81. \end{aligned}$$

d.)

$$\begin{aligned} P(-c < Z < c) &= 0.668 \\ \implies \Phi(c) - \Phi(-c) &= 0.668 \\ \implies \Phi(c) - (1 - \Phi(c)) &= 0.668 \\ \implies \Phi(c) - 1 + \Phi(c) &= 0.668 \\ \implies 2\Phi(c) &= 1.668 \\ \implies \Phi(c) &= 0.834 \\ \implies c &= 0.97. \end{aligned}$$

3. Suppose that the diameter at breast height (in inches) of trees of a certain type is a normally distributed random variable X with mean $\mu = 8.5$ and standard deviation $\sigma = 2.5$. Suppose that one tree of this type is selected at random.

   (a) Find the probability that the diameter of the tree is less than 4.75 inches; i.e., find $P(X < 4.75)$.

   (b) Find the probability that the diameter of the tree is greater than 10 inches; i.e., find $P(X > 10)$.

   (c) Find the probability that the diameter of the tree is between 5 and 15 inches; i.e., find $P(5 < X < 15)$.

   (d) Find the 25$^{\text{th}}$ percentile of the tree diameters; i.e., find the value c so that $P(X \leqslant c) = 0.25$.

   (e) Find the tree diameter for the largest 10% of trees; i.e., find the value c so that $P(X > c) = 0.10$.

   (f) Between what two values are the middle 90% of tree diameters? That is, find the two values L and U so that $P(L < X < U) = 0.90$.

   (g) If two trees are selected independently of each other, what is the probability that both of them are greater than 10 inches?

   (h) If three trees are selected independently of each other, what is the probability that at least one of them has a diameter less than 10 inches?

Remark. When $X \sim N(\mu, \sigma^2)$, probabilities involving X are computed by "standardizing." The *standardized variable* is $(X - \mu)/\sigma$. Subtracting μ shifts the mean from μ to zero, and then dividing by σ scales the variable so that the standard deviation is 1 rather than σ.

If X has a normal distribution with mean μ and standard deviation σ, then

$$Z = \frac{X - \mu}{\sigma}$$

has a standard normal distribution. Thus

$$\begin{aligned} P(a \leqslant X \leqslant b) &= P\left(\frac{a-\mu}{\sigma} \leqslant Z \leqslant \frac{b-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$

$$P(X \leqslant a) = \Phi\left(\frac{a-\mu}{\sigma}\right) \qquad P(X \geqslant b) = 1 - \Phi\left(\frac{b-\mu}{\sigma}\right)$$

The $(100p)^{\text{th}}$ percentile of a normal distribution with mean μ and standard deviation σ is easily related to the $(100p)^{\text{th}}$ percentile of the standard normal distribution.

$$(100p)^{\text{th}} \text{ percentile for normal } (\mu, \sigma) = \mu + \left[(100p)^{\text{th}} \text{ for standard normal}\right] \cdot \sigma$$

Another way of saying this is that if z is the desired percentile for the standard normal distribution, then the desired percentile for the normal (μ, σ) distribution is z standard deviations from μ. ☺

We have $X \sim N(8.5, 2.5^2)$

a.)

$$\begin{aligned} P(X < 4.75) &= P\left(Z < \frac{4.75 - 8.5}{2.5}\right) = P(Z < -1.5) \\ &= \Phi(-1.5) = 1 - \Phi(1.5) \\ &= 1 - 0.9332 = 0.0668. \end{aligned}$$

b.)

$$\begin{aligned} P(X > 10) &= P\left(Z > \frac{10 - 8.5}{2.5}\right) \\ &= P(Z > 0.6) = 1 - \Phi(0.6) \\ &= 1 - 0.7257 = 0.2743. \end{aligned}$$

c.)

$$\begin{aligned} P(5 < X < 15) &= P\left(\frac{5 - 8.5}{2.5} < Z < \frac{15 - 8.5}{2.5}\right) \\ &= P(-1.4 < Z < 2.6) = \Phi(2.6) - \Phi(-1.4) \\ &= \Phi(2.6) - (1 - \Phi(1.4)) \\ &= 0.9953 - 0.0808 = 0.9145. \end{aligned}$$

d.)

$$\begin{aligned} P(X \leqslant c) &= 0.25 \\ \implies P\left(Z \leqslant \frac{c - 8.5}{2.5}\right) &= 0.25 \\ \implies \frac{c - 8.5}{2.5} &= -0.67 \\ \implies c &= -0.67 \cdot 2.5 + 8.5 \\ \implies c &= 6.825. \end{aligned}$$

e.)

$$\begin{aligned} P(X > c) &= 0.1 \\ \implies P\left(Z > \frac{c - 8.5}{2.5}\right) &= 0.1 \\ \implies 1 - P\left(Z < \frac{c - 8.5}{2.5}\right) &= 0.1 \\ \implies P\left(Z < \frac{c - 8.5}{2.5}\right) &= 0.9 \\ \implies \frac{c - 8.5}{2.5} &= 1.28 \\ \implies c &= 11.7. \end{aligned}$$

f.)

$$\begin{aligned} P(L < X < U) &= 0.9 \\ \implies P(X < U) - P(X < L) &= 0.9 \\ \implies P\left(Z < \frac{U - 8.5}{2.5}\right) - P\left(Z < \frac{L - 8.5}{2.5}\right) &= 0.9. \end{aligned}$$

We need to find the 5% and 95% percentile such that $P(X < U) = 0.95$ and $P(X < L) = 0.05$. Thus,

$$\begin{aligned}
P\left(Z < \frac{U - 8.5}{2.5}\right) &= 0.05 \\
\implies \frac{U - 8.5}{2.5} &= -1.64 \\
\implies U &= 4.4.
\end{aligned}$$

$$\begin{aligned}
P\left(Z < \frac{L - 8.5}{2.5}\right) &= 0.95 \\
\implies \frac{L - 8.5}{2.5} &= 1.65 \\
\implies L &= 12.625.
\end{aligned}$$

Thus,

$$P(4.4 < X < 12.625) = 0.9.$$

g.) First, we find

$$\begin{aligned}
P(X > 10) = P\left(Z > \frac{10 - 8.5}{2.5}\right) = 1 - P\left(Z < \frac{10 - 8.5}{2.5}\right) \\
= 1 - \Phi(0.6) = 1 - 0.7257 = 0.2743.
\end{aligned}$$

That is, the probability of a selected tree having a diameter greater than 10 inches is 0.2743. The probability that two trees selected independently of each other having a diameter greater than 10 inches is

$$0.2743^2 = 0.0752.$$

h.) The probability that out of three independently selected trees at least one of them has a diameter less than 10 is the complement of the probability that all three have a diameter greater than 10. That is,

$$1 - (0.2743)^3 = 0.9794.$$

4. Suppose that 80% of all drivers in a certain region regularly wear a seat belt. Let X be the number of drivers out of a random sample of 500 drivers who regularly wear a seat belt. Find the (approximate) probability of each of the following events.

(a) $P(X \leqslant 380)$

(b) $P(390 \leqslant X \leqslant 410)$

Remark. Let X be a binomial rv based on n trials with success probability p. Then if the binomial probability histogram is not too skewed, X has approximately a normal distribution with $\mu = np$ and $\sigma = \sqrt{npq}$. In particular, for $x =$ a possible value of X,

$$\begin{aligned} P(X \leqslant x) = B(x, n, p) &\approx \text{(area under the normal curve to the left of } x + 0.5) \\ &= \Phi\left(\frac{x + 0.5 - np}{\sqrt{npq}}\right) \end{aligned}$$

In practice, the approximation is adequate provided that both $np \geqslant 10$ and $nq \geqslant 10$, since there is then enough symmetry in the underlying binomial distribution. ☺

First, we check the conditions

1. $np \geqslant 10$
2. $n(1-p) \geqslant 10$

$$\begin{aligned} np &= 500(0.8) = 400 \geqslant 10 \\ n(1-p) &= 500(0.2) = 100 \geqslant 10. \end{aligned}$$

Thus, this binomial experiment can be approximated by the normal distribution. We have $\mu = np = 400$ and $\sigma = \sqrt{np(1-p)} = \sqrt{80} = 8.9443$

a.)

$$\begin{aligned} P(X \leqslant 380) = B(380; 500, 0.8) &\approx \Phi\left(\frac{380 + 0.5 - 400}{8.9443}\right) \\ &= \Phi(-2.1802) = 1 - \Phi(2.1802) \\ &= 1 - 0.9854 = 0.0146. \end{aligned}$$

b.)

$$\begin{aligned} P(390 \leqslant X \leqslant 410) &= \Phi\left(\frac{410 + 0.5 - 400}{8.9443}\right) - \Phi\left(\frac{390 - 0.5 - 400}{8.9443}\right) \\ &= \Phi(1.17) - \Phi(-1.17) \\ &= \Phi(1.17) - (1 - \Phi(1.17)) \\ &= 0.8790 - 0.121 = 0.758. \end{aligned}$$

5. Quality audit records are kept on the numbers of major and minor failures that occur to a certain type of circuit pack used during the burn-in period of large electronic switching devices. Let

$$X = \text{the number of major failures}$$

$$Y = \text{the number of minor failures}$$

Suppose that the random variables X and Y can be described, at least approximately, by the joint probability mass function given below.

$p(x,y)$	$y=0$	$y=1$	$y=2$	$y=3$	$y=4$
$x=0$	0.15	0.10	0.10	0.10	0.05
$x=1$	0.05	0.08	0.14	0.08	0.05
$x=2$	0.01	0.01	0.02	0.03	0.03

(a) Find the probability that a randomly selected circuit pack will have 1 major and 2 minor failures.

(b) Find $P(X \leqslant 1 \text{ and } Y \leqslant 1)$.

(c) Find the probability that a randomly selected circuit pack will have fewer major failures than minor failures; i.e., find $P(X < Y)$.

(d) Suppose that demerits are assigned to a circuit pack according to the formula $D = 5X + Y$. Find the probability that a randomly selected circuit pack scores 7 or fewer demerits; i.e., find $P(D \leqslant 7)$.

(e) (i) Give the marginal probability mass function of X.
(ii) Find the mean value of X; i.e., find $E(X)$.
(iii) Find the variance of X.

(f) (i) Give the marginal probability mass function of Y.
(ii) Find the mean value of Y; i.e., find $E(Y)$.
(iii) Find the variance of Y.

(g) (i) Find $E(XY)$.
(ii) Find the expected number of demerits for a circuit pack; i.e., find $E(D)$.

(h) Are X and Y independent random variables? Clearly answer yes or no and explain why or why not.

(i) Find $\text{Cov}(X,Y)$.

(j) Find $\text{Corr}(X,Y)$.

a.)

$$P(X = 1, Y = 2) = 0.14.$$

b.)

$$\begin{aligned} P(X \leqslant 1 \text{ and } Y \leqslant 1) &= \sum_{x=0}^{1}\sum_{y=0} 1\ p(x,y) \\ &= p(0,0) + p(0,1) + p(1,0) + p(1,1) \\ &= 0.15 + 0.10 + 0.05 + 0.08 = 0.38. \end{aligned}$$

c.)

$$
\begin{aligned}
P(X<Y) &= \sum_{(x,y):}\sum_{x<y} p(x,y) \\
&= p(0,1)+p(0,2)+p(0,3)+p(0,4) \\
&+ p(1,2)+p(1,3)+p(1,4)+p(2,3)+p(2,4) \\
&= 0.1+0.1+0.1+0.05+0.14+0.08+0.05+0.03+0.03 \\
&= 0.68.
\end{aligned}
$$

d.) First, we check all pairs to see if they satisfy D.

$$
\begin{aligned}
D(0,0) &= 5(0)+0 = 0 \leqslant 7 \\
D(0,1) &= 5(0)+1 = 1 \leqslant 7 \\
D(0,2) &= 5(0)+2 = 2 \leqslant 7 \\
D(0,3) &= 5(0)+3 = 3 \leqslant 7 \\
D(0,4) &= 5(0)+4 = 4 \leqslant 7 \\
D(1,0) &= 5(1)+0 = 5 \leqslant 7 \\
D(1,1) &= 5(1)+1 = 6 \leqslant 7 \\
D(1,2) &= 5(1)+2 = 7 \leqslant 7 \\
D(1,3) &= 5(1)+3 = 8 \nleqslant 7.
\end{aligned}
$$

These are the points we are interested in. Thus,

$$
\begin{aligned}
P(D\leqslant 7) &= \sum_{(x,y):}\sum_{D(x,y)\leqslant 7} p(x,y) \\
&= p(0,0)+p(0,1)+p(0,2)+p(0,3) \\
&+ p(0,4)+p(1,0)+p(1,1)+p(1,2) \\
&= 0.15+3(0.10)+2(0.05)+0.08+0.14 \\
&= 0.77.
\end{aligned}
$$

e.i)

$$
\begin{aligned}
p_X(x) &= \sum_y p(x,y) \\
p_X(0) &= \sum_y p(0,y) = 0.15+0.1+0.1+0.1+0.05 = 0.5 \\
p_X(1) &= \sum_y p(1,y) = 0.05+0.08+0.14+0.08+0.05 = 0.4 \\
p_X(2) &= \sum_y p(2,y) = 0.01+0.01+0.02+0.03+0.03 = 0.1.
\end{aligned}
$$

Thus,

$$
p_x(x) = \begin{cases} 0.5 & \text{if } x=0 \\ 0.4 & \text{if } x=1 \\ 0.1 & \text{if } x=2 \\ 0 & \text{otherwise} \end{cases}.
$$

e.ii) We use the marginal probabilities for x found above to compute the expected value $E(X)$

$$
\begin{aligned}
E(X) &= \sum_x x \cdot p_X(x) \\
&= 0(0.5) + 1(0.4) + 2(0.1) \\
&= 0.6.
\end{aligned}
$$

e.iii) The variance is given by $E(X^2) - [E(X)]^2$.

$$
\begin{aligned}
E(X^2) &= \sum_x x^2 p_X(x) = 0^2(0.5) + 1^2(0.4) + 2^2(0.1) = 0.8 \\
[E(X)]^2 &= 0.6^2 = 0.36 \\
\therefore V(X) &= 0.8 - 0.36 = 0.44.
\end{aligned}
$$

f.i) The marginal pmf of Y is given by

$$
\begin{aligned}
p_Y(Y) &= \sum_x p(x, y) \\
p_Y(0) &= \sum_x p(x, 0) = 0.15 + 0.05 + 0.01 = 0.21 \\
p_Y(1) &= \sum_x p(x, 1) = 0.10 + 0.08 + 0.01 = 0.19 \\
p_Y(2) &= \sum_x p(x, 2) = 0.10 + 0.14 + 0.02 = 0.26 \\
p_Y(3) &= \sum_x p(x, 3) = 0.10 + 0.08 + 0.03 = 0.21 \\
p_Y(4) &= \sum_x p(x, 4) = 0.05 + 0.05 + 0.03 = 0.13.
\end{aligned}
$$

Thus,

$$
p_Y(y) = \begin{cases} 0.21 & \text{if } y = 0, 30.19 \\ \text{if } y = 10.26 & \text{if } y = 20.13 \\ \text{if } y = 40 & \text{otherwise} \end{cases} .
$$

f.ii) The expected value $E(Y)$ is given by

$$
\begin{aligned}
E(Y) &= \sum_y y \cdot p_Y(y) \\
&= 0(0.21) + 1(0.19) + 2(0.26) + 3(0.21) + 4(0.13) = 1.86.
\end{aligned}
$$

f.iii) The variance is given by

$$
\begin{aligned}
E(Y^2) &= \sum_y y^2 \cdot p_Y(y) = 0^2(0.21) + 1^2(0.19) + 2^2(0.26) + 3^2(0.21) + 4^2(0.13) = 5.2 \\
[E(Y)]^2 &= 1.86^2 = 3.4596 \\
V(Y) &= E(Y^2) - [E(Y)]^2 = 5.2 - 3.4596 = 1.7404.
\end{aligned}
$$

g.i) $E(XY)$ is given by

$$\begin{aligned}
E(XY) = \sum_{(x,y)} xy \cdot p(x,y) &= \sum_x \sum_y xy \cdot p(x,y) \\
&= 0 \cdot 0 \cdot \cdot 0.15 + 0 \cdot 1 \cdot 0.1 + 0 \cdot 2 \cdot 0.1 \\
&\quad + 0 \cdot 3 \cdot 0.1 + 0 \cdot 4 \cdot 0.05 + 1 \cdot 0 \cdot 0.05 \\
&\quad + 1 \cdot 1 \cdot 0.08 + 1 \cdot 2 \cdot 0.14 + 1 \cdot 3 \cdot 0.08 \\
&\quad + 2 \cdot 2 \cdot 0.02 + 2 \cdot 3 \cdot 0.03 + 2 \cdot 4 \cdot 0.03 \\
&\quad + 1 \cdot 4 \cdot 0.05 + 2 \cdot 0 \cdot 0.01 + 2 \cdot 1 \cdot 0.01 \\
&= 1.32.
\end{aligned}$$

g.ii) The expected value $E(D)$ is given by

$$\begin{aligned}
E(D) &= \sum_x \sum_y D(x,y) \cdot p(x,y) \\
&= 0 \cdot 0.15 + 1 \cdot 0.1 + 2 \cdot 0.1 + 3 \cdot 0.1 \cdot 4 + 0.05 \\
&\quad + 5 \cdot 0.05 + 6 \cdot 0.08 + 7 \cdot 0.14 + 8 \cdot 0.08 + 9 \cdot 0.05 \\
&\quad + 10 \cdot 0.01 + 11 \cdot 0.01 + 12 \cdot 0.02 + 13 \cdot 0.03 + 14 \cdot 0.03 \\
&= 4.86.
\end{aligned}$$

h.) X and Y are independent iff $p_X(x) \cdot p_Y(y) = p(x,y)\ \forall (x,y)$

$$p(0,0) = 0.15 \neq 0.5(0.21) = 0.105.$$

Since this does not hold, we conclude X and Y are not independent.

i.) The covariance is given by

$$Cov(X,Y) = E\left[(x-\mu_x)(y-\mu_y)\right] = \sum_x \sum_y (x-\mu_x)(y-\mu_y)p(x,y).$$

Also, by the shortcut formula $Cov(X,Y) = E(XY) - \mu_x \mu_y$, we have

$$\begin{aligned}
Cov(X,Y) &= 1.32 - 0.6 \cdot 1.86 \\
&= 0.204.
\end{aligned}$$

j.) The Correlation Coefficient $\rho_{X,Y}$ is given by $\frac{Cov(X,Y)}{\sigma_X \sigma_Y}$. Thus, we have

$$\begin{aligned}
\rho_{X,Y} &= \frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}} \\
&= \frac{0.204}{\sqrt{0.44}\sqrt{1.7404}} \\
&= 0.2331.
\end{aligned}$$