Homework #2 Solutions

Derek Sonderegger

1. The population distribution of blood donors in the United States based on race/ethnicity and blood type as reported by the American Red Cross is given here:

	O	A	B	AB	Total
White	36%	32.2%	8.8%	3.2%	
Black	7%	2.9%	2.5%	0.5%	
Asian	1.7%	1.2%	1%	0.3%	
Other	1.5%	0.8%	0.3%	0.1%	
					100%

Notice that the numbers given in the table sum to 100%, so the data presented are the probability of a particular ethnicity and blood type.

a) Fill in the column and row totals.

	O	A	B	AB	Total
White	36%	32.2%	8.8%	3.2%	**80.2%**
Black	7%	2.9%	2.5%	0.5%	**12.9%**
Asian	1.7%	1.2%	1%	0.3%	**4.2%**
Other	1.5%	0.8%	0.3%	0.1%	**2.7%**
	46.2%	**37.1%**	**12.6%**	**4.1%**	100%

b) What is the probability that a randomly selected donor will be Asian and have Type O blood? That is to say, given a donor is randomly selected from the list of all donors, what is the probability that the selected donor will Asian with Type O?

$$P(Asian \cap TypeO) = 0.017$$

c) What is the probability that a randomly selected donor is white? That is to say, given a donor is randomly selected from the list of all donors, what is the probability that the selected donor is white?

$$P(White) = 0.802$$

d) What is the probability that a randomly selected donor has Type A blood? That is to say, given a donor is selected from the list of all donors, what is the probability that the selected donor has Type A blood?

$$P(TypeA) = 0.371$$

e) What is the probability that a white donor will have Type A blood? That is to say, given a donor is randomly selected from the list of all the white donors, what is the probability that the selected donor has Type A blood? (Notice we already know the donor is white because we restricted ourselves to that subset!)

$$P(TypeA \middle| White) = \frac{P(TypeA \cap White}{P(White)} = \frac{0.322}{0.802} = 0.401$$

f) Is blood type and ethnicity independent? Justify your response mathematically using your responses from the previous answers.

They are not independent because $0.401 = P(A|White) \neq P(A) = 0.371$.

2. For each of the following, mark if it is Continuous or Discrete.

a) ______________ Milliliters of tea drunk per day.

Continuous - While I probably wouldn't do this, there is nothing that prevents me from drinking 56.332 mL of tea.

b) ______________ Different brands of soda drunk over the course of a year.

Because the idea of 1/2 a brand doesn't make sense, I can only have integer numbers for this response. So discrete.

c) ______________ Number of days per week that you are on-campus for any amount of time.

Notice this is on campus for any amount of time, so it can only be the integers 0 to 7... thus this is discrete.

d) ______________ Number of grizzly bears individuals genetically identified from a grid of hair traps in Glacier National Park.

Here we can only observe integer numbers of individuals so again this is discrete.

3. For each scenario, state whether the event should be modeled via a binomial or Poisson distribution.

a) ______________ Number of M&Ms I eat per hour while grading homework.

This is the number of things per unit time... so a rate and there isn't really an upper bound on this number. So Poisson.

b) ______________ The number of mornings in the coming 7 days that I change my son's first diaper of the day.

This is an integer and it can only be a max of 7 days. This feels like the number of successes out of 7 trials. So Binomial.

c) ______________ The number of Manzanita bushes per 100 meters of trail.

Again this is a rate... ie the number of things per unit distance. So Poisson.

4. During a road bike race, there is always a chance a crash will occur. Suppose the probability that at least one crash will occur in any race I'm in is $\pi = 0.2$ and that races are independent.

a) What is the probability that the next two races I'm in will both have crashes?

$$\begin{aligned} P\left(C_1 \text{ and } C_2\right) &= P\left(C_1\right)P\left(C_2\right) \quad \text{by independence} \\ &= (0.2)\,(0.2) \\ &= 0.04 \end{aligned}$$

b) What is the probability that neither of my next two races will have a crash?

$$\begin{aligned} P\left(\bar{C}_1 \text{ and } \bar{C}_2\right) &= P\left(\bar{C}_1\right)P\left(\bar{C}_2\right) \quad \text{by independence} \\ &= (0.8)\,(0.8) \\ &= 0.64 \end{aligned}$$

c) What is the probability that at least one of the next two races have a crash?

$$\begin{aligned}P(\text{at least 1 crash}) &= 1 - P(\text{no crashes}) \\ &= 1 - P(\bar{C}_1 \text{ and } \bar{C}_2) \\ &= 1 - 0.64 \\ &= 0.36\end{aligned}$$

5. My cats suffer from gastric distress due to eating house plants and the number of vomits per week that I have to clean up follows a Poisson distribution with rate $\lambda = 1.2$ pukes per week.

So the random variable V is the number of vomits I have to clean up and

$$V \sim \textit{Poisson}\,(\lambda = 1.2 \text{ per week})$$

a) What is the probability that I don't have to clean up any vomits this coming week?

I could calculate this by hand...

$$\begin{aligned}P(V=0) &= \frac{\lambda^0 e^{-\lambda}}{\lambda!} \\ &= \frac{1 \cdot e^{-1.2}}{0!} \\ &= e^{-1.2} \\ &= 0.3011\end{aligned}$$

Or I could do it in R

```
r    dpois(0, lambda=1.2)
## [1] 0.3011942
```

b) What is the probability that I must clean up 1 or more vomits?

One or more is just the complement of 0, so I could calculate

$$P(V \geq 1) = 1 - P(V \leq 0) = 1 - P(V = 0) = 1 - 0.302 = 0.698$$

```
r    1 - dpois(0, lambda=1.2)
## [1] 0.6988058
```

c) If I wanted to measure this process with a rate per day, what rate should I use?

If we were interested in the number of vomits per day... then because I expect $\lambda = 1.2$ per week, then I would expect

$$\lambda = \frac{1.2 \text{ vomit}}{\text{week}} \cdot \frac{1 \text{ week}}{7 \text{ days}} = \frac{0.1714 \text{ vomit}}{\text{day}}$$

6. Suppose that the number of runners I see on a morning walk on the trails near my house has the following distribution (Notice I've never seen four or more runners on a morning walk):

y	0	1	2	3	4+
Probabilty	0.45	0.25	0.20		0.0

a) What is the probability that I see 3 runners on a morning walk? _Because the probabilities need to sum to 1, then the 3 must have a probability of 0.10.

b) What is the expected number of runners that I will encounter?

$$\begin{aligned}\mu = E(Y) &= \sum_{y=0}^{3} y \cdot P(Y = y) \\ &= 0 \cdot (0.45) + 1 \cdot (0.25) + 2 \cdot (0.20) + 3 \cdot (0.10) \\ &= 0 + 0.25 + 0.40 + 0.30 \\ &= 0.95\end{aligned}$$

c) What is the variance of the number of runners that I will encounter?

$$\begin{aligned}Var(Y) &= \sum_{y=0}^{3} (y-\mu)^2 \cdot P(Y = y) \\ &= (0 - .95)^2 \cdot (0.45) + (1 - .95)^2 \cdot (0.25) + (2 - .95)^2 \cdot (0.20) + (3 - .95)^2 \cdot (0.10) \\ &= (-.95)^2 \cdot (0.45) + (.05)^2 \cdot (0.25) + (1.05)^2 \cdot (0.20) + (2.05)^2 \cdot (0.10) \\ &= 1.0475\end{aligned}$$

I'll admit that I double checked the calculation using R

```
x <- 0:3
p <- c(0.45, 0.25, 0.20, 0.1)
mu <- sum( x*p )
mu
```

```
## [1] 0.95
```

```
(x-mu)^2
```

```
## [1] 0.9025 0.0025 1.1025 4.2025
```

```
sum( (x-mu)^2 * p )
```

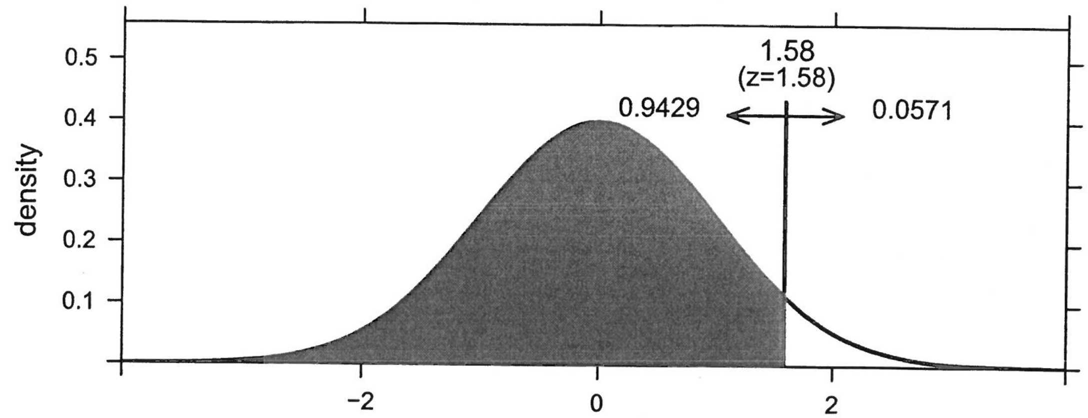
```
## [1] 1.0475
```

7. If $Z \sim N\left(\mu = 0, \sigma^2 = 1\right)$, find the following probabilities:

a) $P(Z < 1.58) = 0.943$

```
mosaic::xpnorm(1.58)
```

```
##
## If X ~ N(0, 1), then
##
##  P(X <= 1.58) = P(Z <= 1.58) = 0.9429466
##  P(X >  1.58) = P(Z >  1.58) = 0.05705343
```



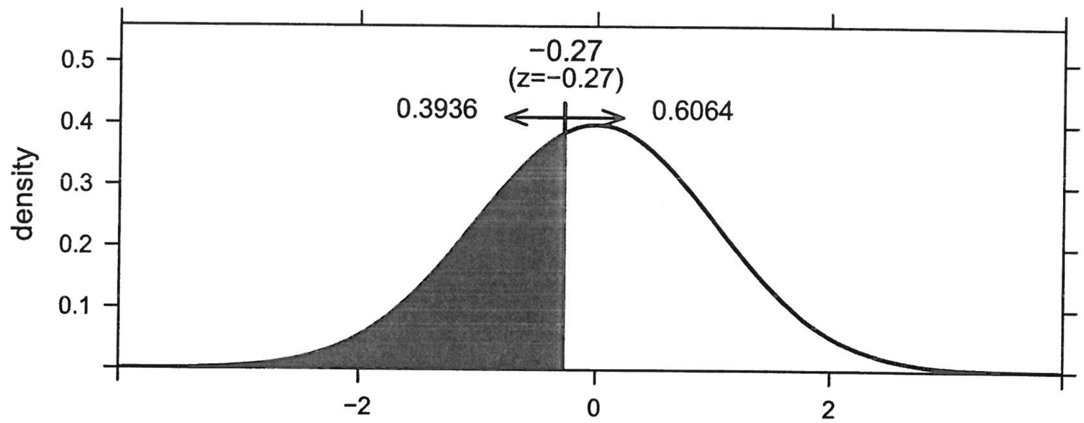

```
## [1] 0.9429466
```

b) $P(Z = 1.58) = 0$ because the probability of any particular result is 0.

c) $P(Z > -.27) = 0.3935$

```
mosaic::xpnorm(-0.27)
```

```
##
## If X ~ N(0, 1), then
##
##  P(X <= -0.27) = P(Z <= -0.27) = 0.3935801
##  P(X >  -0.27) = P(Z >  -0.27) = 0.6064199
```

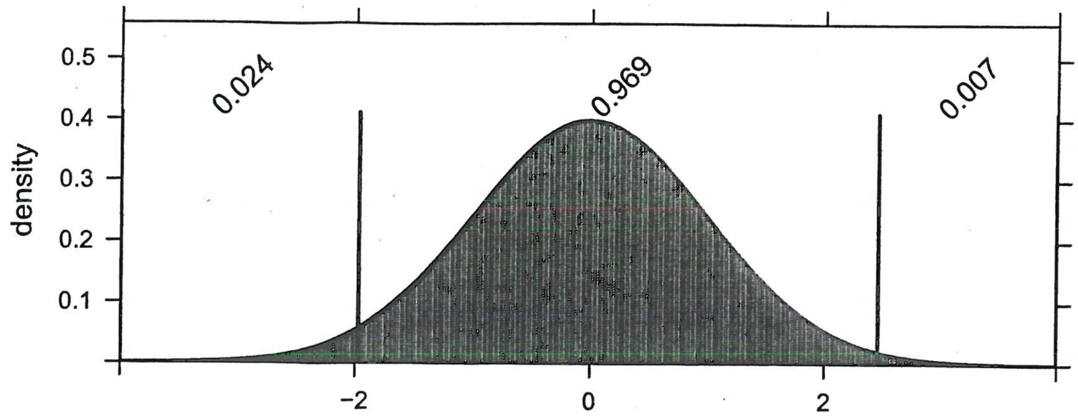



```
## [1] 0.3935801
```

d) $P(-1.97 < Z < 2.46) = P(Z < 2.46) - P(Z < -1.97) = 0.9931 - 0.0244 = 0.9687$

```
mosaic::xpnorm(c(-1.97, 2.46))
```

```
##
## If X ~ N(0, 1), then
##
## P(X <= -1.97) = P(Z <= -1.97) = 0.02441919
##     P(X <=  2.46) = P(Z <=  2.46) = 0.99305315
## P(X >  -1.97) = P(Z >  -1.97) = 0.975580815
##     P(X >   2.46) = P(Z >   2.46) = 0.006946851
```




```
## [1] 0.02441919 0.99305315
```