PubH 7401: Elements of Biostatistical Inference I
Homework 6, due Thursday October 23

HINT: For problems 2–5, 7, and 9 use formulas in your text/notes. These should be relatively quick!

1. On Homework 3 problem 1, find $E(X)$ for problem 1 (health care expenditures). Contrast this to the median and mode of X.
2. On Homework 3 problem 2, find how many we'd expect to have diabetes, $E(X)$, and a measure of spread $\text{sd}(X)$ for X assuming a binomial distribution.
3. On Homework 3 problem 4, find $E(X)$ and $\text{sd}(X)$ in months. Also find the interquartile range of X, $\text{IQR}(X)$.
4. On Homework 3 problem 5, find the expected number of armadillos killed over three days. How does this number compare to the median?
5. On Homework 3 problem 6, find $E(X)$ and $\text{var}(X)$.
6. Let (X, Y) be jointly distributed with pdf

$$f(x,y) = \left\{ \begin{array}{cc} x+y & \text{for } \begin{array}{c} 0 \le x \le 1 \\ 0 \le y \le 1 \end{array} \\ 0 & \text{otherwise} \end{array} \right\}.$$

   (a) Find $E(XY)$, $E(X)$, and $E(Y)$. You can use results from Homework 5 prob. 3.
   (b) Find $\text{var}(X)$, and $\text{var}(Y)$.
   (c) Use (a) to find $\text{cov}(X, Y)$.
   (d) Use (b) and (c) to find ρ_{XY}.
   (e) Use propositions in your notes/text to find $E(X + Y)$ and $\text{var}(X + Y)$.

7. Let X_1 and X_2 be independent random variables, each having distribution $f(x) = \frac{8}{3}x^3 \exp(-2x)$ for $x \ge 0$. See problem 4 on homework 5.
   (a) What is the mean and variance of each random variable: $E(X_i)$ and $\text{var}(X_i)$?
   (b) What is $\text{cov}(X_1, X_2)$? What is the correlation between these two variables?

8. Let (X, Y) be jointly distributed with the pmf

x	y	$p(x,y)$
-1	-1	0.2
-1	1	0.1
0	-1	0.1
0	1	0.3
1	-1	0.2
1	1	0.1

   (a) Are X and Y independent? Why or why not?

   (b) Find $\text{cov}(X, Y)$ and ρ_{XY}.

9. Let $X \sim N(10, 4)$ independent of $Y \sim \text{gamma}(5, 0.1)$. Find the following numbers:

   (a) $E(3X - 2Y + 2)$ and $\text{var}(3X - 2Y + 2)$.

   (b) ρ_{XY}.

10. Chapter 4, problem 14 (p. 168).

   (a) **Brute force**: Do the problem as stated in the text using calculus. Note that you can get $\text{var}(X)$ from $E(X)$ and $E(X^2)$.

   (b) **Clever solution**: What kind of distribution is this? Use the formulas for $E(X)$ and $\text{var}(X)$; do they agree with part (a)? They should!

   (c) **Lazy solution**: Verify that $F(x) = x^2$ and $F^{-1}(p) = \sqrt{p}$ for X with $f(x) = 2x$ on $0 \le x \le 1$. In `R`, use the inverse cdf method to simulate

$$X_1, \dots, X_{100000} \overset{iid}{\sim} f(x) = 2x, \quad 0 \le x \le 1.$$

   From these simulated values, report the sample mean and sample variance. Do they agree with the theoretical values you found? Please print out `R` code and output and turn in with your homework.

11. Chapter 4, problem 31 (p. 169). Show all work.