Problem Set 2: Statistics

(Due Friday 2/24/2017 at 5pm)

You will need the following for problems 1 and 2:

In lecture, we talked about the mean of a distribution of data and its standard deviation. Another very important notion is that of an **expected value**. An expected value is the mean of a random variable. In other words, it is what we expect to get as a long-term average value of a numerical **random variable** X. The expected value of X is denoted $E[X]$.

For example, let's say you're playing a game where you roll a die once and you receive an amount of dollars equal to the number of dots on the face you rolled. The expected value of money that you will earn is:

$$\frac{1+2+3+4+5+6}{6} = \$3.5. \tag{1}$$

Now suppose that the numbers $\{1, 2, 3, 4, 5, 6\}$ do not occur with equal probability, but have a probability distribution $\{\frac{1}{21}, \frac{2}{21}, \frac{3}{21}, \frac{4}{21}, \frac{5}{21}, \frac{6}{21}\}$ respectively. Then, the expected number of dollars that you will receive if you roll this die once is:

$$\left(\frac{1}{21} \times 1\right) + \left(\frac{2}{21} \times 2\right) + \left(\frac{3}{21} \times 3\right) + \left(\frac{4}{21} \times 4\right) + \left(\frac{5}{21} \times 5\right) + \left(\frac{6}{21} \times 6\right) = \$4.33. \tag{2}$$

So, in general, the expected value of a random variable X, where X can take the values $\{x_1, x_2, \ldots, x_n\}$ with probabilities $\{p_1, p_2, \ldots, p_n\}$ is:

$$E[X] = (p_1 \times x_1) + (p_2 \times x_2) + \cdots + (p_n \times x_n). \tag{3}$$

Problem 1

A manufacturer is sending 10 boxes out for shipment today. Unfortunately, some of the boxes have defective items

Box #	1	2	3	4	5	6	7	8	9	10	
#of defective items	0	0	1	0	2	2	0	0	1	3	

(4)
(5)

(a) One of these boxes is to be selected at random for shipment to a particular customer. Let X be the number of defective items in the selected box. What is the probability distribution of X.

(b) What is the expected value of defective items?

(c) Another manufacturer is known to have X^2 defective items in each of the boxes numbered 1 to 10. If this manufacturer sends out a randomly selected box, what is the expected number of defective items the customer will receive?

Problem 2

You toss a fair coin 10 times. Let X = number of heads obtained in 10 tosses.

(a) What is the probability distribution of X?

(b) What is the expected value of X?

(c) Now suppose the coin isn't fair and has a higher probability of obtaining Heads, with $P(H) = 0.7$ and $P(T) = 0.3$. What is the expected value of Heads for this coin when it is tossed 10 times? To calculate this, perform steps (a) and (b).

(d) For the coin in (c), can you guess the expected number of Heads in 1000 tosses? (Doing the calculation will probably take all week).

(e) In general, if $P(H) = p$ and $P(T) = 1 - p$, what is the expected number of Heads in n tosses?

Problem 3

Women of the Dinaric Alps are considered among the tallest in the world with an average height of 171.1cm (5ft 7.5in), while the average height of an American woman is 162.9cm (5ft 4in). Suppose that both heights are approximately normally distributed with a standard deviation of 6cm.

(a) What is the probability that a randomly selected woman from the Dinaric Apls region is shorter than the average American woman?

(b) What is the probability that a randomly selected woman from the US is taller than the average Dinaric Alps woman?

(c) Suppose that you have a representative group of women from the Dinaric Alps and from the US mixed together. If you select a woman at random (there is a 50-50 chance of selecting a woman from the Dinaric Alps region or from the US), what is the probability that this woman is taller than 170cm?

Problem 4

This problem is selected from Chapter 4 in: Devore, J. L., *Probability and Statistics*, 9th Edition (2016)

Vehicle speed on a particular bridge in China can be modeled as normally distributed ("Fatigue Reliability Assessment for Long-Span Bridges under Combined Dynamic Loads from Winds and Vehicles," J. of Bridge Engr., 2013: 735–747).

(a) If 5% of all vehicles travel less than 39.12 mi/h and 10% travel more than 73.24 mi/h, what are the mean and standard deviation of vehicle speed?
Hint: A drawing will help a great deal. Also, use the table (attached at the end). If you don't find the exact number you're looking for, just use the one closest to it.

(b) What is the probability that a randomly selected vehicle's speed is between 50 and 65 mi/h?
Hint: $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$.

(c) What is the probability that a randomly selected vehicle's speed exceeds the speed limit of 70 mi/h?

Problem 5

This problem is selected from Chapter 5 in: Devore, J. L., *Probability and Statistics*, 9th Edition (2016)

Let X denote the courtship time for a randomly selected female–male pair of mating scorpion flies (time from the beginning of interaction until mating). Suppose the mean value of X is 120 min and the standard deviation of X is 110 min (suggested by data in the article "Should I Stay or Should I Go? Condition- and Status-Dependent Courtship Decisions in the Scorpion Fly Panorpa Cognate" (Animal Behavior, 2009: 491–497)).

(a) Is it plausible that X is normally distributed?

(b) For a random sample of 50 such pairs, what is the (approximate) probability that the sample mean courtship time is between 100 min and 125 min?

(c) For a random sample of 50 such pairs, what is the (approximate) probability that the total courtship time exceeds 150 min?

(d) Could the probability requested in (b) be calculated from the given information if the sample size were 15 rather than 50? Explain.

Problem 6

A community has two hospitals. Hospital A is a large medical center, while Hospital B is a more fashionable and much more expensive hospital where most patients are wealthy. An article in the local paper claims that a higher percentage of surgery patients die at Hospital A than at Hospital B, and deplores the fact that people who are less well off are disadvantaged. It also recommends to the people in the community that if they can afford it, they should choose to have their surgery in Hospital B rather than A. A more detailed look at the number of surgery patients in the last few months at both Hospitals, taking into account also whether the incoming patients were in good or poor health, shows the following:

	Hospital A		Hospital B	
	Good Health	Poor Health	Good Health	Poor Health
Died	4	57	5	8
Survived	559	1422	585	196

(a) What are the percentages of patients admitted for surgery who are in bad health prior to the operation?

(b) What are the total percentages of patients who died?

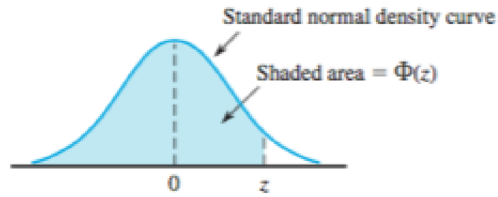
(c) What are the percentages of patients in previously good health who died?

(d) What are the percentages of patients in previously poor health who died?

(e) Try describing this paradox in your own words in one paragraph. Don't just repeat the numbers: give a clear explanation of what is happening in these statistics.

Table A.3 Standard Normal Curve Areas

$\Phi(z) = P(Z \le z)$



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0038
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3482
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Table A.3 Standard Normal Curve Areas *(cont.)* $\Phi(z) = P(Z \leq z)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998