University of California, Davis
Department of Biomedical Engineering

Fall 2013	**Probability and Statistics**	**BIM 105**
David M. Rocke	**for Biomedical Engineers**	**October 31, 2013**

Midterm Examination

NAME ______________________________

1. We have alkaline phosphatase measurements for 20 breast cancer patients at diagnosis as given in the table below. Compute the five number summary and the inner fences (that are used in constructing the boxplot). Are there any apparent outliers? Explain.

92	102	115	128	145	146	150	150	153	161
173	175	178	180	182	191	191	213	228	230

Solution (20 points)

- $n = 20$ so the rank of the median is $(n + 1)/2 = 10.5$. The median is $(161 + 173)/2 = 167$.
- The ranks of the first and third quartiles are $21/4 = 5.25$ and $3 \times 21/4 = 15.75$, so the first and third quartiles are $(145 + 146)/2 = 145.5$ and $(182 + 910)/2 = 186.5$.
- The five-number summary is $(92, 145.5, 167, 186.5, 230)$.
- The IQR is $186.5 - 145.5 = 41$. The inner fences are 1.5 IQRs out from the quartiles, so are $145.5 - 41 \times 1.5 = 84$ and $186.5 + 41 \times 1.5 = 248$.
- Since all the observations are within the inner fences, there are no apparent outliers.

2. Given the following data on x and y,

   (a) Compute the mean, variance, and standard deviation of x and y.

   (b) Compute the correlation coefficient.

   (c) Find the least-squares line.

   (d) Find the predicted value of y and the residual when $x = 4$.

x	3	7	2	4
y	8	12	4	12

Solution (20 points)

	x	y	$x-\bar{x}$	$(x-\bar{x})^2$	$y-\bar{y}$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
	3	8	-1	1	-1	1	1
	7	12	3	9	3	9	9
	2	4	-2	4	-5	25	10
	4	12	0	0	3	9	0
Sum	16	36	0	14	0	44	20

(a) $\bar{x} = 16/4 = 4$, $s_x^2 = 14/3 = 4.67$, $s_x = \sqrt{14/3} = 2.16$
$\bar{y} = 36/4 = 9$, $s_y^2 = 44/3 = 14.67$, $s_y = \sqrt{44/3} = 3.83$

(b) $\rho_{xy} = \frac{20}{\sqrt{14 \times 44}} = \frac{20}{\sqrt{616}} = \frac{20}{24.8} = 0.806$

(c) $\hat{\beta}_1 = \frac{20}{14} = \frac{10}{7} = 1.429$
$\hat{\beta}_0 = 9 - \frac{10}{7}4 = \frac{23}{7} = 3.286$
The least squares line is $\hat{y} = 3.286 + 1.429x$.

(d) $\hat{y}(4) = \frac{23}{7} + \frac{10}{7}4 = \frac{63}{7} = 9$,
residual $= 12 - 9 = 3$

3. A population of devices is from either of two batches, A and B. Some items are defective and some are not. Suppose that the probability that a randomly chosen device is from batch A is $P(\text{A}) = 0.30$ and the probability that the device is defective is $P(\text{D}) = 0.10$. Suppose also that the probability that a randomly chosen device is from batch B and not defective is $P(\text{B}\cap\text{ND}) = 0.65$.

   (a) Find $P(\text{B})$, $P(\text{ND})$, $P(\text{A} \cap \text{D})$, $P(\text{A} \cap \text{ND})$, and $P(\text{B} \cap \text{D})$. This might be easiest if you make a two-way table.

   (b) Are the events A and D independent?

   (c) Find $P(\text{D}|\text{B})$

   (d) Find any two events that are mutually exclusive.

   Solution (20 points)

	D	ND	
A	0.05	0.25	**0.30**
B	0.05	**0.65**	0.70
	0.10	0.90	1.00

   (a) 0.70, 0.90, 0.05, 0.25, 0.05.

   (b) No. $P(A)P(D) = (0.30)(0.10) = 0.03 \neq P(A \cap D) = 0.05$.

   (c) $P(D|B) = 0.05/0.70 = 1/14 = 0.0714$.

   (d) For example, A and B, D and ND, $A \cap D$ and $A \cap ND$, etc.

4. The useful life X of a knee replacement has a mean of 12 years and a standard deviation of 10 years. When you need it, use the attached copy of the necessary part of Table A.2. You don't need to interpolate.

   (a) If the useful life was normally distributed, what proportion of devices would have a useful life of longer than 20 years? What could you say about this probability if you did not know that the useful life was normally distributed?

   (b) If data were collected on 200 joint replacements what would be the mean and standard deviation of the average useful life $\bar{X}$?

   (c) What is the probability that $\bar{X} > 14$? Does this depend on the assumption of the normality of X? Explain.

Solution (20 points)

(a) $P(X > 20) = P(Z > (20 - 12)/10) = P(Z > 0.8) = 1 - 0.7881 = 0.2119$.
You cannot say much without knowing the specific distribution family. In fact, for this specific question, the probability can be anything between 0 and 1.

(b) $E(\bar{X}) = 12$, $V(\bar{X}) = 100/200 = 0.5$, $SD(\bar{X}) = \sqrt{0.5} = 0.7071$.
You need to assume independence. You do not need normality.

(c) $P(\bar{X} > 14) = P(Z > 2/0.7071) = P(Z > 2.83) = 1 - 0.9977 = 0.0023$.
You do not need normality of X, because n is large and the Central Limit Theorem says that $\bar{X}$ is then normally distributed even if X is not.

5. A random variable X has PDF

$$f(x) = \begin{cases} k(1-x^2) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Find k.

(b) Find the CDF, $F(x)$. What is the probability that $X < 0.5$?

(c) Find the mean, variance, and standard deviation of X.

Solution (20 points)

(a) Since the PDF must integrate to 1, we have
$\int_0^1 k(1-x^2)dx = k(x - x^3/3)|_0^1 = k(1-1/3) = 2k/3$ so $k = 3/2$.

(b) The CDF is

$$F(x) = \begin{cases} 0 & x < 0 \\ (3/2)(x - x^3/3) & 0 \le x \le 1 \\ 1 & x > 1 \end{cases}$$

$P(X < 0.5) = (3/2)(0.5 - 0.5^3/3) = 0.6875,$

(c) $\mu_X = E(X) = \int_0^1 (3/2)x(1-x^2)dx = (3/2)(x^2/2 - x^4/4)|_0^1$
$= (3/2)(1/2 - 1/4) = 3/8 = 0.375.$
$E(X^2) = \int_0^1 (3/2)x^2(1-x^2)dx = (3/2)(x^3/3 - x^5/5)|_0^1$
$= (3/2)(1/3 - 1/5) = 1/5 = 0.2.$
$\sigma_X^2 = V(X) = E(X^2) - \mu_X^2 = 1/5 - (3/8)^2 = 0.059375.$
$\sigma_X = \sqrt{0.059375} = 0.2437$

6. The number of defects in a long fiber-optic cable is Poisson distributed with a rate of 1 defect per 800m.

   (a) Find the probability of at least one defect in 2000m of cable.

   (b) Each defect in this type of cable costs \$3 to repair. In addition, the cable housing independently has Poisson distributed defects at 1 defect per 500m, and each of these cost \$2 to repair. If we have 2000m of cable and cable housing, let X be the random variable of the cost of repair for the cable and the cable housing. Find the mean, variance, and standard deviation of X.

   Solution (20 points)

   (a) With 2000m of cable, the mean number of defects is $2000/800 = 2.5$, so $\lambda = 2.5$. The chance of no defect is
   $P(X \geq 1) = 1 - f(0) = 1 - \frac{\lambda^0}{0!}e^{-\lambda} = 1 - e^{-2.5} = 1 - 0.08208 = 0.9179$

   (b) X_1 has mean 2.5 and variance 2.5. X_2 has mean 4 and variance 4. The cost is $Y = 3X_1 + 2X_2$.
   $E(Y) = (3)(2.5) + (2)(4) = 7.5 + 8 = \$15.50$
   $V(Y) = (3)^2(2.5) + (2)^2(4) = 22.5 + 16 = 38.5$
   $SD(Y) = \sqrt{38.5} = \$6.20$

7. Transformation of fibroblasts into iPSC stem cells is obtained by inducing expression of certain genes. This succeeds 4% of the time, and suppose that this is independent from cell to cell.

   (a) If 200 cells are processed, what is the mean and standard deviation of the number of cells X successfully transformed?

   (b) What is the chance that $X = 0$? That $X = 1$? That $X = 2$?

   Solution (20 points)

   (a) $X = \text{Bin}(200, 0.04)$
   $E(X) = (200)(0.04) = 8$
   $\text{Var(X)} = (200)(0.04)(0.96) = 7.68$
   $SD(X) = \sqrt{7.68} = 2.77$

   (b) $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$
   $P(X = 0) = \binom{200}{0}(0.04)^0(0.96)^{200} = (0.96)^{200} = 0.00028$
   $P(X = 1) = \binom{200}{1}(0.04)^1(0.96)^{199} = (200)(0.04)(0.96)^{199} = 0.00237$
   $P(X = 2) = \binom{200}{2}(0.04)^2(0.96)^{198} = (19900)(0.04)^2(0.96)^{198} = 0.00983$