Probability and Statistics, Spring 2006, BCC CUNY

Instructor: Natalia Novak

Preview Test 1 (chapters 1-4)

1. Use a random-number table to simulate the outcomes of tossing a quarter 23 times. Assume the quarter is balanced (i.e. fair).

   Solution: The possible outcomes of tossing a quarter once are: Heads (**H**) and Tails (**T**). Since we will use the random-number **Table 1** for simulation, let's assume that *even* number means *Heads* and *odd* number means *Tails*. Take 23 digits from the table (row 6 block 2): *51709 94456 48396 73780 06436 86641* (all 0's are crossed out, since they are neither even nor odd)
   And a possible outcome of tossing a quarter 23 times is: *T T T T T H H T H H H T T H T T T H H H T H H.*

2. Suppose you have been walking for 20 minutes each day for 2 weeks. And for each day you recorded the distance you covered in 20 minutes. Distance is measured in miles. Make a time plot.

3. Driving under the influence of alcohol (DUI) is a serious offense. The following data give the ages of a random sample of 50 drivers arrested while driving under influence of alcohol. This distribution is based on the age distribution of DUI arrests given in the Statistical Abstract of the United States (112th Edition). (see your handouts for the values)

   a) Make stem-and-leaf display of the age distribution: since all the the numbers are two-digit numbers, let a first digit be a stem and a second digit be a leaf.

Stem	Leaf	Stem	Leaf
5	3 stands for 53 years	5	3 stands for 53 years
1	6, 8	1	6, 8
2	6, 2, 2, 1, 6, 7, 4, 6, 7, 2, 0, 5, 9, 3, 4, 1, 7	2	0, 1, 1, 2, 2, 2, 3, 4, 4, 5, 6, 6, 6, 7, 7, 7, 9
3	3, 0, 6, 4, 1, 8, 1, 5, 7, 2, 9, 0, 4, 5	3	0, 0, 1, 1, 2, 3, 4, 4, 5, 5, 6, 7, 8, 9
4	6, 1, 7, 3, 0, 9, 9, 0, 5, 7	4	0, 0, 1, 3, 5, 6, 7, 7, 9, 9
5	5, 8, 3, 6, 1	5	1, 3, 5, 6, 8
6	3, 4	6	3, 4

Conclusion: it seems that people in their 20s-30s tend to drink and drive more than others.

b) Make a histogram showing class boundaries (don't forget to make a frequency table first). Use seven classes.

1) Let's find class width: $\frac{64-16}{7} = \frac{48}{7} \approx 6.86$. Thus *class width* $= 7$.

2) Frequency table:

Class Limits (Lower-Upper)	Class boundaries (Lower-Upper)	Midpoint	Frequency	Relative Frequency
16-22	15.5-22.5	$\frac{16+22}{2} = 19$	8	$\frac{8}{50} = 0.16$
23-29	22.5-29.5	$\frac{23+29}{2} = 26$	11	$\frac{11}{50} = 0.22$
30-36	29.5-36.5	$\frac{30+36}{2} = 33$	11	$\frac{11}{50} = 0.22$
37-43	36.5-43.5	$\frac{37+43}{2} = 40$	7	$\frac{7}{50} = 0.14$
44-50	43.5-50.5	$\frac{44+50}{2} = 47$	6	$\frac{6}{50} = 0.12$
51-57	50.5-57.5	$\frac{51+57}{2} = 54$	4	$\frac{4}{50} = 0.08$
58-64	57.5-64.5	$\frac{58+64}{2} = 61$	3	$\frac{3}{50} = 0.06$

4. Radon is a gas emitted from the ground that can collect in houses and buildings. At certain levels it can cause lung cancer. Radon concentrations are measured in picocuries per litter (pCi/L). A radon level of 4 pCi/L is considered "'acceptable"'. Radon levels in a house vary from week to week. In one house, a sample of 8 weeks had the following readings for radon level (in pCi/L): 1.9 2.8 5.7 4.2 1.9 8.6 3.9 7.2

a) Find the mean, median and mode

Let's order data first: 1.9 1.9 2.8 3.9 4.2 5.7 7.2 8.6

$mode = 1.9, \qquad median = \frac{3.9+4.2}{2} = 4.05$

$mean(\bar{x}) = \frac{1.9+1.9+2.8+3.9+4.2+5.7+7.2+8.6}{8} = 4.525$

b) Find the sample standard deviation, coefficient of variation and the range.

$range = 8.6 - 1.9 = 6.7$, then

x	x^2
1.9	3.61
1.9	3.61
2.8	7.84
3.9	15.21
4.2	17.64
5.7	32.49
7.2	51.84
8.6	73.96
$\Sigma x = 36.2$	$\Sigma x^2 = 206.2$

Sample variance $s^2 = \frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1} = \frac{206.2 - \frac{36.2^2}{8}}{7} = \frac{42.395}{7} \approx 6.056$, and

standard deviation $s = \sqrt{42.3957} \approx 2.46$

$CV = \frac{s}{\bar{x}} \cdot 100 = \frac{2.46}{4.525} \cdot 100 \approx 54.36$

c) Find quartiles Q_1, Q_2, Q_3, make box-and-whiskers plot. Find IQR.

$Q_2 = median = 4.05, \quad Q_1 = \frac{1.9+2.8}{2} = 2.35, \quad Q_3 = \frac{5.7+7.2}{2} = 6.45,$

$IQR = Q_3 - Q_1 = 6.45 - 2.35 = 4.1$

5. Modern medical practice tells us not to encourage babies to become too fat. Is there a positive correlation between the weight x of a 1-year-old baby and the weight y of a mature adult (30 years old)? (see the data values in the handout)

   (a) Draw a scatter diagram for the data.

(b) r is positive

(c,d) Find r, coefficient of determination, and the equation of the least-squares line

x	y	x^2	y^2	xy
21	125	441	15625	2625
25	125	625	15625	3125
23	120	529	14400	2760
24	125	576	15625	3000
20	130	400	16900	2600
15	120	225	14400	1800
25	145	625	21025	3625
21	130	441	16900	2730
17	130	289	16900	2210
24	130	576	16900	3120
26	130	676	16900	3380
22	140	484	19600	3080
18	110	324	12100	1980
19	115	361	13225	2185
$\Sigma x = 300$	$\Sigma y = 1775$	$\Sigma x^2 = 6572$	$\Sigma y^2 = 226125$	$\Sigma xy = 38220$

$r = \frac{n\Sigma xy - \Sigma x \Sigma y}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}} = \frac{14 \cdot 38220 - 300 \cdot 1775}{\sqrt{14 \cdot 6572 - 300^2}\sqrt{14 \cdot 226125 - 1775^2}} \approx 0.468$

$r^2 \approx 0.219$, $\bar{x} = \frac{300}{14} \approx 21.43$, $\bar{y} = \frac{1775}{14} \approx 126.79$,

$b = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2} = \frac{14 \cdot 38220 - 300 \cdot 1775}{14 \cdot 6572 - 300^2} = \frac{2580}{2008} \approx 1.285$

$a = \bar{y} - b \cdot \bar{x} = \frac{\Sigma y - b\Sigma x}{n} = \frac{1775 - \frac{2580}{2008} \cdot 300}{14} = 99.253$,

$\hat{y} = 99.253 + 1.285x$ - the equation of the least-squares line.

(f) If a female baby weights 20 lb at 1 year, what would you predict she would weight at 30 years of age?

for 20 lb the predicted weight is $99.253 + 1.285 \cdot 20 = 124.953$, so at the age of 30 years the predicted weight is about 125 pounds.