Problem Set #1 - Solution (Fall '04)

2.54 a. The sample mean is:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{5+3+4+\ldots+3}{20} = \frac{80}{20} = 4$$

The sample median is found by finding the average of the 10th and 11th observations once the data are arranged in order. The data arranged in order are:

1 1 1 1 1 2 2 3 3 3 4 4 4 5 5 5 6 7 9 13

The 10th and 11th observations are 3 and 4. The average of these two numbers (median) is:

$$\text{median} = \frac{3+4}{2} = \frac{7}{2} = 3.5$$

The mode is the observation appearing the most. For this data set, the mode is 1, which appears 5 times.

b. Eliminating the largest number which is 13 results in the following:

The sample mean is:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{5+3+4+\ldots+3}{19} = \frac{67}{19} = 3.53$$

The sample median is found by finding the middle observation once the data are arranged in order. The data arranged in order are:

1 1 1 1 1 2 2 3 3 3 4 4 4 5 5 5 6 7 9

The 10th observation is 3. The median is 3

The mode is the observations appearing the most. For this data set, the mode is 1, which appears 5 times.

By dropping the largest number, the mean is reduced from 4 to 3.53. The median is reduced from 3.5 to 3. There is no effect on the mode.

c. The data arranged in order are:

1 1 1 1 1 2 2 3 3 3 4 4 4 5 5 5 6 7 9 13

If we drop the lowest 2 and largest 2 observations we are left with:

1 1 1 2 2 3 3 3 4 4 4 5 5 5 6 7

The sample 10% trimmed mean is:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1+1+1+\ldots+7}{16} = \frac{56}{16} = 3.5$$

The advantage of the trimmed mean over the regular mean is that very large and very small numbers that could greatly affect the mean have been eliminated.

2.62 a. Range = 3 – 0 = 3

$$s^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1} = \frac{15 - \frac{7^2}{5}}{5-1} = 1.3 \qquad s = \sqrt{1.3} = 1.1402$$

b. After adding 3 to each of the data points,

Range = 6 – 3 = 3

$$s^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1} = \frac{102 - \frac{22^2}{5}}{5-1} = 1.3 \qquad s = \sqrt{1.3} = 1.1402$$

c. After subtracting 4 from each of the data points,

Range = –1 – (–4) = 3

$$s^2 = \frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1} = \frac{39 - \frac{(-13)^2}{5}}{5-1} = 1.3 \qquad s = \sqrt{1.3} = 1.1402$$

d. The range, variance, and standard deviation remain the same when any number is added to or subtracted from each measurement in the data set.

2.76 a. From the information given, we have $\bar{x} = 375$ and $s = 25$. From Chebyshev's Rule, we know that at least three-fourths of the measurements are within the interval:

$\bar{x} \pm 2s$, or (325, 425)

Thus, at most one-fourth of the measurements exceed 425. In other words, more than 425 vehicles used the intersection on at most 25% of the days.

b. According to the Empirical Rule, approximately 95% of the measurements are within the interval:

$\bar{x} \pm 2s$, or (325, 425)

This leaves approximately 5% of the measurements to lie outside the interval. Because of the symmetry of a mound-shaped distribution, approximately 2.5% of these will lie below 325, and the remaining 2.5% will lie above 425. Thus, on approximately 2.5% of the days, more than 425 vehicles used the intersection.

2.78 a. Since the sample mean (18.2) is larger than the sample median (15), it indicates that the distribution of years is skewed to the right. In addition, the maximum number of years is 50 and the minimum is 2. If the distribution were symmetric, the mean and median should be about halfway between these two numbers. Halfway between the maximum and minimum values is 26, which is much larger than either the mean or the median.

b. The standard deviation can be estimated by the range divided by either 4 or 6. For this distribution, the range is:

Range = Largest – smallest = 50 – 2 = 48.

Dividing the range by 4, we get an estimate of the standard deviation to be 48/4 = 12.

Dividing the range by 6, we get an estimate of the standard deviation to be 48/6 = 8.

Thus, the standard deviation should be somewhere between 8 and 12. For this problem, the standard deviation is $s = 10.64$. This value falls in the estimated range of 8 to 12.

c. First, we calculate the number of standard deviations from the mean the value of 40 years is. To do this, we first subtract the mean and then divide by the value of the standard deviation.

Number of standard deviations is $\frac{40-\bar{x}}{s}=\frac{40-18.2}{10.64}=2.05\approx 2$

Using Chebyshev's Rule, we know that at most $1/k^2$ or $1/2^2 = 1/4$ of the data will be more than 2 standard deviations from the mean. Thus, this would indicate that at most 25% of the Generation Xers responded with 40 years or more.

Next, we calculate the number of standard deviations from the mean the value of 8 years is.

Number of standard deviations is $\frac{8-\bar{x}}{s}=\frac{8-18.2}{10.64}=-.96\approx -1$

Using Chebyshev's Rule, we get no information about the data within 1 standard deviation of the mean. However, we know the median (15) is more than 8. By definition, 50% of the data are larger than the median. Thus, at least 50% of the Generation Xers responded with 8 years or more. No additional information can be obtained with the information given.

2.96 a. From the problem, $\mu = 2.7$ and $\sigma = .5$

$$z=\frac{x-\mu}{\sigma}\Rightarrow z\sigma = x-\mu \Rightarrow x=\mu+z\sigma$$

For $z = 2.0$, $x = 2.7 + 2.0(.5) = 3.7$

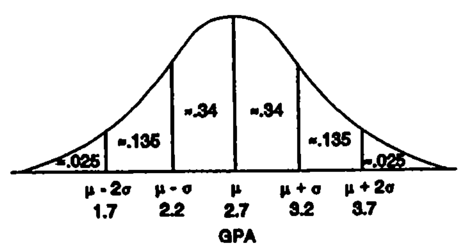
For $z = -1.0$, $x = 2.7 - 1.0(.5) = 2.2$

For $z = .5$, $x = 2.7 + .5(.5) = 2.95$

For $z = -2.5$, $x = 2.7 - 2.5(.5) = 1.45$

b. For $z = -1.6$, $x = 2.7 - 1.6(.5) = 1.9$

c. If we assume the distribution of GPAs is approximately mound-shaped, we can use the Empirical Rule.



From the Empirical Rule, we know that ≈.025 or ≈2.5% of the students will have GPAs above 3.7 (with $z = 2$). Thus, the GPA corresponding to summa cum laude (top 2.5%) will be greater than 3.7 ($z > 2$).

We know that ≈.16 or 16% of the students will have GPAs above 3.2 ($z = 1$). Thus, the limit on GPAs for cum laude (top 16%) will be greater than 3.2 ($z > 1$).

We must assume the distribution is mound-shaped.

2.97 a. The 10th percentile is the score that has at least 10% of the observations less than it. If we arrange the data in order from the smallest to the largest, the 10th percentile score will be the .10(75) = 7.5 or 8th observation. When the data are arranged in order, the 8th observation is 0. Thus, the 10th percentile is 0.

b. The 95th percentile is the score that has at least 95% of the observations less than it. If we arrange the data in order from the smallest to the largest, the 95th percentile score will be the .95(75) = 71.25 or 72nd observation. When the data are arranged in order, the 72nd observation is 21. Thus, the 95th percentile is 21.

2.134 a. **Using MINITAB, the stem-and-leaf display is:**

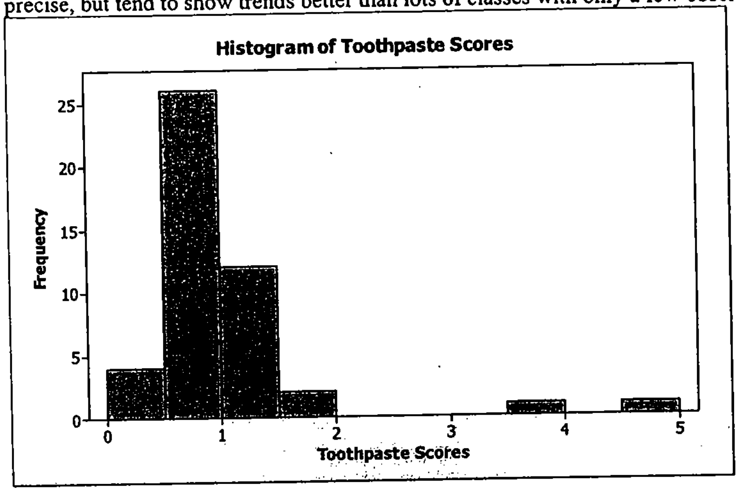
```
Stem-and-leaf of C1
Leaf Unit = 0.10     N = 46

   4     0   3444
 (25)    0   55555555566666667777788889
  16     1   00001122233 4
   4     1   77
   2     2
   2     2
   2     3
   2     3   9
   1     4
   1     4   7
```

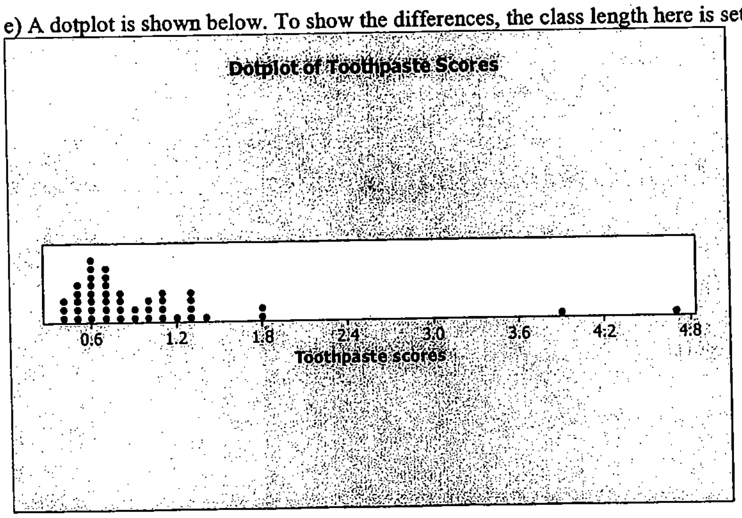
b. **The leaves that represent those brands that carry the American Dental Association seal are circled above.**

c. **It appears that the cost of the brands approved by the ADA tend to have the lower costs. Thirteen of the twenty brands approved by the ADA, or (13/20) × 100% = 65% are less than the median cost.**

d) A histogram is below. Notice that the class length here is .5. There is always a tradeoff between smaller class lengths to show more precision in the breakdown of the data and larger class lengths which aren't as precise, but tend to show trends better than lots of classes with only a few observations in each



e) A dotplot is shown below. To show the differences, the class length here is set at .2.



f) It's a matter of personal preference, of course, but assuming that class lengths are chosen in a histogram to allow for enough precision while still showing trends, a histogram tends to be the most useful graphical display.

2.146

a) With no information, it would seem logical to recommend Clinic A since the "central measure" of weight loss seems to indicate that it is more effective. However, this is with some skepticism, because we know that the mean is more suceptible to outliers than the median. So, it could be that clinic A's mean is higher because of a few extremely high observations even if B's program is more effective for more people.

b) With this information, it seems prudent to recommend Clinic B. At clinic A, mean>median, and so we expect skew to the right (i.e. high outliers). Indeed, the standard deviation is significantly larger at Clinic A. The medians are the same between the two clinics, so it seems like Clinic B is the "safer" (less variable) choice.

c) We'd want to guarantee that the samples were randomly and independently collected.

3.4 a. $\binom{9}{4} = \frac{9!}{4!(9-4)!} = \frac{9\cdot8\cdot7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1}{4\cdot3\cdot2\cdot1\cdot5\cdot4\cdot3\cdot2\cdot1} = 126$

b. $\binom{7}{2} = \frac{7!}{2!(7-2)!} = \frac{7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1}{2\cdot1\cdot5\cdot4\cdot3\cdot2\cdot1} = 21$

c. $\binom{4}{4} = \frac{4!}{4!(4-4)!} = \frac{4\cdot3\cdot2\cdot1}{4\cdot3\cdot2\cdot1\cdot1} = 1$

d. $\binom{5}{0} = \frac{5!}{0!(5-0)!} = \frac{5\cdot4\cdot3\cdot2\cdot1}{1\cdot5\cdot4\cdot3\cdot2\cdot1} = 1$

e. $\binom{6}{5} = \frac{6!}{5!(6-5)!} = \frac{6\cdot5\cdot4\cdot3\cdot2\cdot1}{5\cdot4\cdot3\cdot2\cdot1\cdot1} = 6$

3.16 a. Since order does not matter, the number of different bets would be a combination of 8 things taken 2 at a time.

The number of ways would be

$$\binom{8}{2} = \frac{8!}{2!(8-2)!} = \frac{8\cdot7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1}{2\cdot1\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1} = \frac{40{,}320}{1440} = 28$$

b. If all players are of equal ability, then each of the 28 sample points would be equally likely. Each would have a probability of occurring of 1/28. There is only one sample point with values 2 and 7. Thus, the probability of winning with a bet of 2-7 would by 1/28 or .0357.

3.20 We will denote the five successful utility companies as S_1, S_2, S_3, S_4, and S_5 and the two failing companies as F_1 and F_2. There are

$$\binom{7}{3} = \frac{7!}{3!4!} = \frac{7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1}{3\cdot2\cdot1\cdot4\cdot3\cdot2\cdot1} = 35$$

possible ways to choose three companies from the seven, as shown below:

(S_1, S_2, S_3)	(S_1, S_3, S_4)	(S_1, S_4, S_5)	(S_1, S_5, F_1)
(S_1, S_2, S_4)	(S_1, S_3, S_5)	(S_1, S_4, F_1)	(S_1, S_5, F_2)
(S_1, S_2, S_5)	(S_1, S_3, F_1)	(S_1, S_4, F_2)	
(S_1, S_2, F_1)	(S_1, S_3, F_2)		(S_1, F_1, F_2)
(S_1, S_2, F_2)			

(S_2, S_3, S_4)	(S_2, S_4, S_5)	(S_2, S_5, F_1)	(S_2, F_1, F_2)
(S_2, S_3, S_5)	(S_2, S_4, F_1)	(S_2, S_5, F_2)	
(S_2, S_3, F_1)	(S_2, S_4, F_2)		
(S_2, S_3, F_2)			

(S_3, S_4, S_5)	(S_3, S_5, F_1)	(S_3, F_1, F_2)
(S_3, S_4, F_1)	(S_3, S_5, F_2)	
(S_3, S_4, F_2)		

(S_4, S_5, F_1)	(S_5, F_1, F_2)
(S_4, S_5, F_2)	
(S_4, F_1, F_2)	

a. Each outcome is equally likely, so each sample point has probability 1/35. From the 35 events listed, 10 do not contain F_1 or F_2. Therefore, P(selecting none) = 10/35.

b. From the 35 events listed, 20 contain either F_1 or F_2, but not both. Therefore, P(selecting one) = 20/35.

c. From the 35 events listed, 5 contain both F_1 and F_2. Therefore, P(selecting both) = 5/35.

3.26 a. $P(A^c) = P(E_3) + P(E_6) = .2 + .3 = .5$

b. $P(B^c) = P(E_1) + P(E_7) = .10 + .06 = .16$

c. $P(A^c \cap B) = P(E_3) + P(E_6) = .2 + .3 = .5$

d. $P(A \cup B) = P(E_1) + P(E_2) + P(E_3) + P(E_4) + P(E_5) + P(E_6) + P(E_7)$
$= .10 + .05 + .20 + .20 + .06 + .30 + .06 = .97$

e. $P(A \cap B) = P(E_2) + P(E_4) + P(E_5) = .05 + .20 + .06 = .31$

f. $P(A^c \cup B^c) = P(E_1) + P(E_7) + P(E_3) + P(E_6) = .10 + .06 + .20 + .30 = .66$

g. No. A and B are mutually exclusive if $P(A \cap B) = 0$. Here, $P(A \cap B) = .31$.

3.38 a. Yes, the probabilities in the table sum to 1.

$$.05 + .16 + .05 + .19 + .32 + .05 + .11 + .05 + .02 = 1$$

b. $P(A) = .05 + .16 + .05 = .26$
$P(B) = .05 + .19 + .11 = .35$
$P(C) = .05 + .16 + .19 + .32 = .72$
$P(D) = .05 + .05 + .11 + .05 + .02 = .28$
$P(E) = .05$

c. $P(A \cup B) = .05 + .16 + .05 + .19 + .11 = .56$
$P(A \cap B) = .05$
$P(A \cup C) = .05 + .16 + .05 + .19 + .32 = .77$

d. $P(A^c) = 1 - P(A) = 1 - .26 = .74$

The probability that a managerial prospect is not highly motivated is .74. Only about 1/4 of the prospects are highly motivated.

e. The pairs of events that are mutually exclusive have no sample points in common.

$A \cap B$ contains the event "Prospect places in the high motivation category and in the high talent category." Therefore, A and B are not mutually exclusive.

$A \cap C$ contains the event "Prospect places in the high motivation category and in the medium or high talent category." Therefore, A and C are not mutually exclusive.

$A \cap D$ contains the event "Prospect places in the high motivation category and in the low talent category." Therefore, A and D are not mutually exclusive.

$A \cap E$ contains the event "Prospect places in the high motivation category and in the high talent category." Therefore, A and E are not mutually exclusive.

$B \cap C$ contains the event "Prospect places in the high talent category and in the medium or high motivation category." Therefore, B and C are not mutually exclusive.

$B \cap D$ contains the event "Prospect places in the high talent category and in the low motivation category." Therefore, B and D are not mutually exclusive.

$B \cap E$ contains the event "Prospect places in the high talent category and in the high motivation category." Therefore, B and E are not mutually exclusive.

$C \cap D$ contains no events. Therefore, C and D are mutually exclusive.

$C \cap E$ contains the event "Prospect places in the high talent category and in the high motivation category." Therefore, C and E are not mutually exclusive.

$D \cap E$ contains no events. Therefore, D and E are mutually exclusive.

3.42 a. Since A and B are mutually exclusive events, $P(A \cup B) = P(A) + P(B) = .30 + .55 = .85$

b. Since A and C are mutually exclusive events, $P(A \cap C) = 0$

c. $$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0}{.55} = 0$$

d. Since B and C are mutually exclusive events, $P(B \cup C) = P(B) + P(C) = .55 + .15 = .70$

e. No, B and C cannot be independent events because they are mutually exclusive events.

3.54 Define the following events:

A_i: {ith CEO has bachelors degree}

a. $P(A_1) = \frac{6}{25} = .24$

b. $P(A_5 \mid A_1 \cap A_2 \cap A_3 \cap A_4) = \frac{2}{21} = .095$

3.56 If A and B are independent, then $P(A \cap B) = P(A)P(B)$. For this Exercise,

$P(A) = \frac{1385 + 786}{3934} = \frac{2171}{3934} = .552$, $P(B) = \frac{1385 + 1175}{3934} = \frac{2560}{3934} = .651$, and

$P(A \cap B) = \frac{1385}{3934} = .352$.

$P(A)P(B) = .552(.651) = .359 \neq .352 = P(A \cap B)$. Thus, A and B are not independent.

3.64 Define the following event:

E: {Error produced when dividing}

From the problem, $P(E)$ = 1/9,000,000,000

a. $P(E^c) = 1 - P(E) = 8{,}999{,}999{,}999/9{,}000{,}000{,}000 = .999999999 \approx 1.0000$

b. $P(E^c \cap E^c) = .999999999(.999999999) = .999999999 \approx 1$
(assuming the events are independent)

c. Assuming that all 1,000,000,000 divides are independent, the probability that there will be no errors in 1,000,000,000 divides is:

$$P(E^c \cap E^c \cap E^c \cdots \cap E^c) = [P(E^c)]^{1{,}000{,}000{,}000}$$
$$= (8{,}999{,}999{,}999/9{,}000{,}000{,}000)^{1{,}000{,}000{,}000}$$
$$= .9048$$

d. The event "at least one error in 1 billion divisions" is the complement of the event described in part **c**. Thus, the probability of at least one error in 1 billion divisions is:

$1 - .9048 = .0952$

3.110 a. Define the following events:

A_1: {Component 1 works properly}
A_2: {Component 2 works properly}
B_3: {Component 3 works properly}
B_4: {Component 4 works properly}
A: {Subsystem A works properly}
B: {Subsystem B works properly}

The probability a component fails is .1, so the probability a component works properly is $1 - .1 = .9$.

Subsystem A works properly if both components 1 and 2 work properly.

$$P(A) = P(A_1 \cap A_2) = P(A_1)P(A_2) = .9(.9) = .81$$
(since the components operate independently)

Similarly, $P(B) = P(B_1 \cap B_2) = P(B_1)P(B_2) = .9(.9) = .81$

The system operates properly if either subsystem A or B operates properly.

The probability the system operates properly is:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A)P(B)$$
$$= .81 + .81 - .81(.81) = .9639$$

b. The probability exactly one subsystem fails is:

$$P(A \cap B^c) + P(A^c \cap B) = P(A)P(B^c) + P(A^c)P(B)$$
$$= .81(1 - .81) + (1 - .81).81 = .1539 + .1539 = .3078$$

c. The probability the system fails is the probability that both subsystems fail or:

$$P(A^c \cap B^c) = P(A^c)P(B^c) = (1 - .81)(1 - .81) = .0361$$

d. The system operates correctly 99% of the time means it fails 1% of the time. The probability one subsystem fails is .19. The probability n subsystems fail is $.19^n$. Thus, we must find n such that

$$.19^n \leq .01$$

Thus, $n = 3$.

Problem X1

(a)

$$\frac{\sum_{i=1}^{4}(\sqrt{x_i}+3)}{\sum_{i=1}^{4}2\sqrt{(y_i)^2}} = \frac{(\sqrt{x_1}+3)+(\sqrt{x_2}+3)+(\sqrt{x_3}+3)+(\sqrt{x_4}+3)}{2\sqrt{(y_1)^2}+2\sqrt{(y_2)^2}+2\sqrt{(y_3)^2}+2\sqrt{(y_4)^2}}$$

$$= \frac{(\sqrt{x_1}+\sqrt{x_2}+\sqrt{x_3}+\sqrt{x_4})+(3+3+3+3)}{2\sqrt{(y_1)^2}+2\sqrt{(y_2)^2}+2\sqrt{(y_3)^2}+2\sqrt{(y_4)^2}}$$

$$= \frac{(\sqrt{16}+\sqrt{4}+\sqrt{1}+\sqrt{25})+12}{2\sqrt{(9)^2}+2\sqrt{(0)^2}+2\sqrt{(4)^2}+2\sqrt{(-1)^2}}$$

$$= \frac{(4+2+1+5)+12}{2\left(\sqrt{81}+\sqrt{0}+\sqrt{16}+\sqrt{1}\right)}$$

$$= \frac{24}{2(9+0+4+1)}$$

$$= \frac{24}{28} = \frac{6}{7}$$

$$\approx .857$$

(b)

$$\frac{\sum_{i=2}^{5}(\sqrt{x_i})+3}{2\sqrt{\left(\sum_{i=1}^{4}y_i\right)^2}} = \frac{(\sqrt{x_2}+\sqrt{x_3}+\sqrt{x_4}+\sqrt{x_5})+3}{2\sqrt{(y_1+y_2+y_3+y_4)^2}}$$

$$= \frac{(\sqrt{4}+\sqrt{1}+\sqrt{25}+\sqrt{4})+3}{2\sqrt{(9+0+4+(-1))^2}}$$

$$= \frac{2+1+5+2+3}{2\sqrt{(12)^2}}$$

$$= \frac{13}{2(12)} = \frac{13}{24}$$

$$\approx .542$$

(c)

$$\sum_{i=1}^{4}\frac{(\sqrt{x_i}+3)}{2\sqrt{(y_i)^2}} = \frac{\sqrt{x_1}+3}{2\sqrt{(y_1)^2}}+\frac{\sqrt{x_2}+3}{2\sqrt{(y_2)^2}}+\frac{\sqrt{x_3}+3}{2\sqrt{(y_3)^2}}+\frac{\sqrt{x_4}+3}{2\sqrt{(y_4)^2}}$$

$$= \frac{\sqrt{16}+3}{2\sqrt{(9)^2}}+\frac{\sqrt{4}+3}{2\sqrt{(0)^2}}+\frac{\sqrt{1}+3}{2\sqrt{(4)^2}}+\frac{\sqrt{25}+3}{2\sqrt{(-1)^2}}$$

$$= \frac{4+3}{2\sqrt{(9)^2}}+\frac{2+3}{2\sqrt{(0)^2}}+\frac{1+3}{2\sqrt{(4)^2}}+\frac{5+3}{2\sqrt{(-1)^2}}$$

$$= \frac{7}{2\sqrt{81}}+\frac{5}{2\sqrt{0}}+\frac{4}{2\sqrt{16}}+\frac{8}{2\sqrt{1}}$$

$$= \frac{7}{2*9}+\frac{5}{2*0}+\frac{4}{2*4}+\frac{8}{2*1}$$

Notice that this sum is undefined because of the division by zero in the second term.