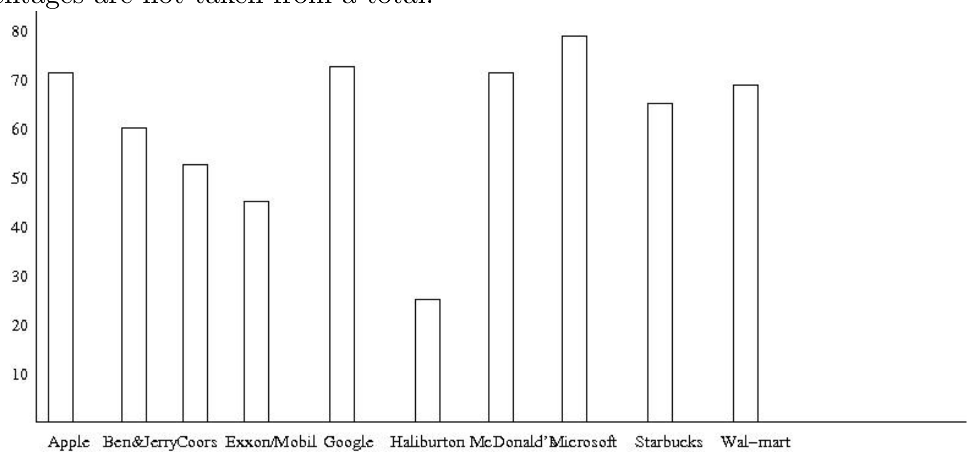
AMS 102: PRACTICE FOR MIDTERM 1

SOLUTIONS

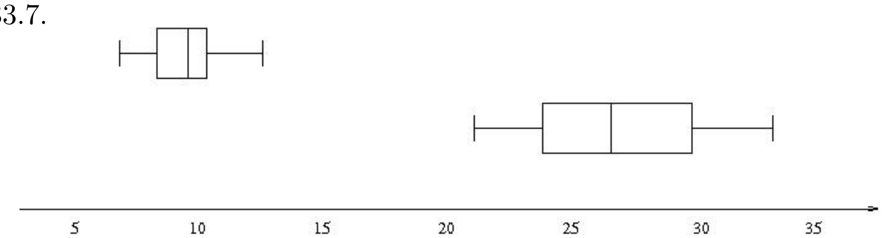
7.3. We have to use a bar graph: the variable is categorical and the percentages are not taken from a total:



7.9. Five-number summaries:

Aleppo pine: Min= 7.2, Q1= 8.6, Median= 9.3, Q3= 10.55, Max= 12.8.

Torrey pine: Min= 21.2, Q1= 23.95, Median= 26.7, Q3= 29.7, Max= 33.7.



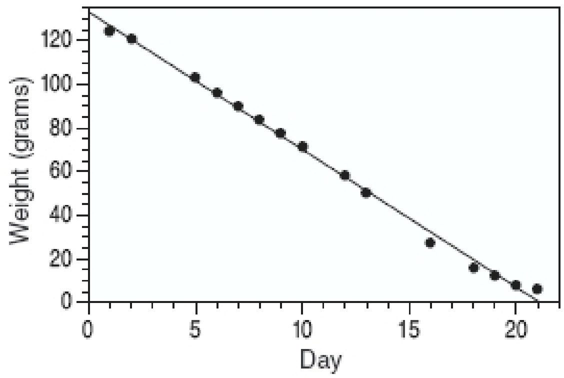
Clearly just by looking at the needles we can distinguish between two types of pine: the longest Aleppo pine needle is shorter than the shortest Torrey pine needle.

7.14. $M - Q1 = 14.9$ pounds and $Q3 - M = 24.1$ pounds, so the weights are skewed to the right. In particular, this makes the mean higher than the median.

7.15. Center 95% lies within two standard deviations from the mean: $9.6 \pm 2(1.6) = 6.4$-12.8 cm.

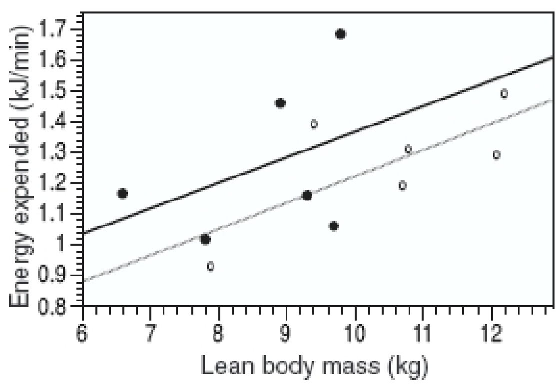
Since $6.4 = \text{mean} - 2$ standard deviations, needles less than 6.4 cm long lie in the bottom $\dfrac{100-95}{2} = 2.5\%$.

7.19. The association is negative (as day increases, weight decreases), so r should be negative. The scatterplot shows a very strong linear relationship, so r should be close to -1.



7.23. (a) For lean monkeys, the mean lean body mass is 8.6833 kg; for the obese monkeys the mean is 10.5167 kg.

(b) On the scatterplot below solid circles are datapoints for lean monkeys; open circles, for obese monkeys. Energy increases at about the same rate for lean and obese monkeys. However, obese monkeys expend less energy overall. Most likely, the excess calories are stored as fat:



(For the record, the lines are $\hat{y} = 0.541 + 0.0826x$ and $\hat{y} = 0.371 + 0.0852x$ for lean and obese monkeys, respectively. But regresssion line could have been drawn by hand here.)

7.29. (a) Fidelity Technology Fund has a larger correlation, and so is more closely tied to the stock market.

(b) No. Correlations describe only relative sizes of variables (above/below average).

7.31. The article is incorrect. A correlation of 0.8 means that the regression line explains about $r^2 = 64\%$ of the variation of European stock prices.

7.32. (a) Explanatory: weeds per meter is explanatory. Response: corn yield.

(b) The stemplots suggest that yields decrease when there are more lamb's-quarter weeds:

0 weeds per meter		1 weed per meter		3 weeds per meter		9 weeds per meter	
14		14		14		14	2
14		14		14		14	
15		15		15	3	15	
15		15	7	15	69	15	
16		16	1	16		16	233
16	57	16	67	16		16	
17	2	17		17		17	
17	7	17		17	6	17	

It appears that greater density of weed diminishes the yield.

(c) Medians: 169.45, 163.65, 157.3, 162.6

Means: 170.2, 162.83, 161.03, 157.6.

Using means instead of medians doesn't affect the conclusion in general. However, the medians do not allow us to conclude that "the more weed, the lower the yield," thus suggesting that more research is needed. The means allow us to skirt the issue.

7.34. (a) The marginal distribution is

Save time	21.20%
Easy	21.20%
Low price	27.72%
Far from store	8.15%
No pressure	7.07%
Other reason	14.67%

(Percentages are computed out of the total of 184 students.)

(b) The conditional distributions in question are

	American students	Asian students
Save time	25.22%	14.49%
Easy	24.35%	15.94%
Low price	14.78%	49.28%
Far from store	9.57%	5.80%
No pressure	8.70%	4.35%
Other reason	17.39%	10.14%

7.41. We need to compare seeds' masses in the two groups. Stemplots will work very well here:

Cicada plants		Control plants
0	1	
	1	3
4	1	445
7	1	77
99	1	89999
111100	2	0111
3333332222	2	2
5544	2	4444445555
7777666	2	66666
999	2	89
110	3	
	3	
5	3	

Overall, the stemplots are similar. We conclude that adding cicadas makes no difference. That is, the data does not support the hypothesis that cicadas make good fertilizer.

7.47. We have to determine how parent opinions about schools differ among these three groups.

Conditional distribution is the way to go:

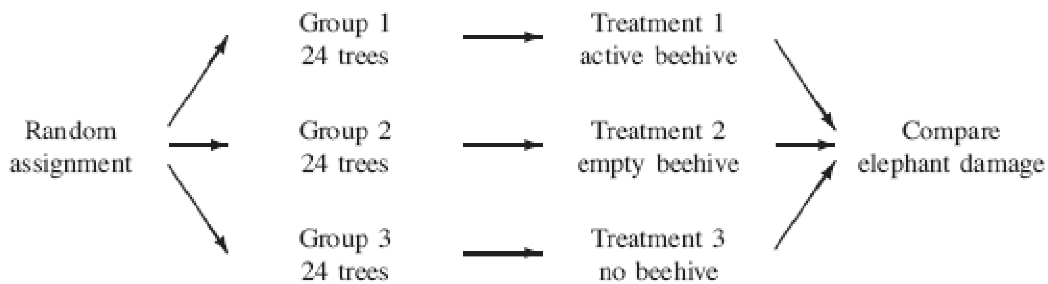
	Black	Hispanic	White
Excellent	5.9%	16.8%	10.9%
Good	34.2%	27.2%	40%
Fair	37.1%	30.2%	29.7%
Poor	11.9%	11.9%	11.9%
Don't know	10.9%	13.9%	6%

We can see that black parents are generally less excited about schools than either hispanic or white parents. (Whether the numbers are statistically significant or not is impossible to determine without knowing more details about the survey.) You may reach other conclusions as well.

16.1. This is an experiment. The participants are directed to do something (and even introduced to the problem beforehand). The response variable is a student's ability to detect fraud. The explanatory variable is alertness.

16.2. The population is 1400 students who participated in counseling. The sample is 200 students who were mailed questionnaires. (That not everybody responded is a typical example of non-response bias.)

16.5. (a) The control group should have 24 trees with no beehives. See diagram below:



(b) Label the trees 01 through 72. From line 137 of Table B, the first four active-hive trees are 53, 64, 56, and 68.

(c) Response variable: elephant damage.

16.9. Only people who watched the program and wanted to take part in the survey participated (volunteer sample). They certainly do not reflect the American adults in general.