Statistics 151 Fall 2006 (Kozdron) Midterm #1 — Solutions

1. (a) If a success is defined as "students passes" then $p = 2/3$ and so the probability that every student passes is $P(120) = \frac{120!}{0!120!}\left(\frac{2}{3}\right)^{120}\left(\frac{1}{3}\right)^{0} = \left(\frac{2}{3}\right)^{120}$.

1. (b) Since there are $n = 120$ students, and $p = 4/6$, we conclude that the number of students expected to pass is $120 \times 2/3 = 80$.

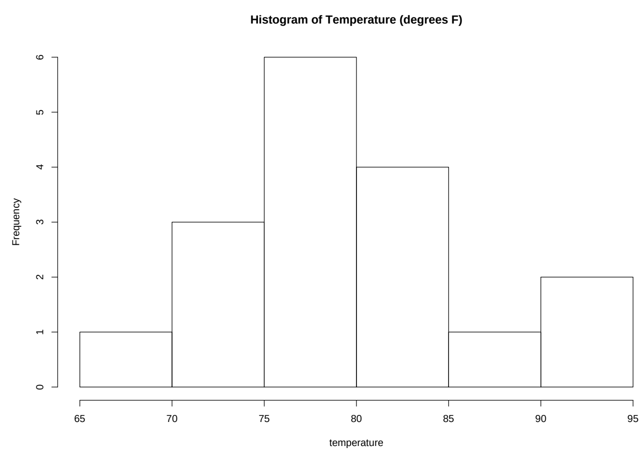
1. (c) The probability that at most 3 students fail is $P(120 \text{ pass}) + P(119 \text{ pass}) + P(118 \text{ pass}) + P(117 \text{ pass})$

$$= \frac{120!}{0!120!}\left(\frac{2}{3}\right)^{120}\left(\frac{1}{3}\right)^{0} + \frac{120!}{1!119!}\left(\frac{2}{3}\right)^{119}\left(\frac{1}{3}\right)^{1} + \frac{120!}{2!118!}\left(\frac{2}{3}\right)^{118}\left(\frac{1}{3}\right)^{2} + \frac{120!}{3!117!}\left(\frac{2}{3}\right)^{117}\left(\frac{1}{3}\right)^{3}.$$

2. (a) The stem-and-leaf plot is

5	
6	9
7	3 4 4 6 7 8 8 8 9
8	2 2 3 4 9
9	1 3

2. (b) The histogram is



2. (c) The mean temperature is

$$\bar{x} = \frac{74+78+82+83+78+73+84+89+91+78+77+82+69+93+76+74+79}{17}$$
$$= \frac{1360}{17} = 80°F.$$

2. (d) The variance is

$$s^2 = \frac{(74-80)^2 + (78-80)^2 + \cdots + (79-80)^2}{17-1}$$
$$= \frac{36+4+4+9+4+49+16+81+121+4+9+4+121+169+16+36+1}{16}$$
$$= \frac{684}{16} = 42.75$$

and so the standard deviation is $s = \sqrt{42.75} = 6.54°F$.

2. (e) Clearly, the minimum is 69, and the maximum is 93. The median is 78 since it is the 9th largest value. Therefore, $75 = (74 + 76)/2$, the median of the numbers strictly less than 78, is the first quartile, and $83.5 = (83 + 84)/2$, the median of numbers strictly larger than 78, is the third quartile. Hence, the five-number summary is $(69, 75, 78, 83.5, 93)$.

2. (f) Judging by the stem-and-leaf plot, it appears that the data is skewed, in which case the median is the most appropriate measure of centre. However, from the histogram, it appears that the data is reasonably symmetric, in which case both the mean and median are equivalent as measures of centre.

2. (g) Notice that a linear change to every data point causes the same change to be done to the mean. Therefore, $\overline{X} = \frac{9}{5} \cdot (80 - 32) = 86.4°C$. *Note that there is a major problem here. The formula for fahrenheit is incorrect, and the 5/9 should actually be a 9/5. In this case,* $\overline{X} = \frac{5}{9} \cdot (80 - 32) = 26.7°C$ *which is perfectly reasonable.*

3. (a) Let X denote the BMD value, which is $N(0.85, 0.11^2)$. Since the mean is 0.85 and the standard deviation is 0.11, we know that 2.5 standard deviations is 0.275. Hence, someone has osteoporosis if her BMD test value is below $0.85 - 0.275 = 0.575$. Thus, Jessica does not have osteoporosis since her value of 0.67 is greater than 0.575.

3. (b) Similar to **(a)**, Allison does not have osteoporosis since her value of 1.13 is greater than 0.575.

3. (c) From Table E, we find that the middle 75% of standard normal data lies between $z = -1.15$ and $z = 1.15$. Since X is $N(0.85, 0.11^2)$ and $X = \sigma z + \mu$, we conclude that X lies between $X = 0.11 \times -1.15 + 0.85 = 0.7235$ and $X = 0.11 \times 1.15 + 0.85 = 0.9765$.

3. (d) We find $P(0.70 < X < 1.05) = P\left(\frac{0.70-0.85}{0.11} < \frac{X-0.85}{0.11} < \frac{1.05-0.85}{0.11}\right) = P(-1.36 < z < 1.82) = 0.4131 + 0.4656 = 0.8787$ where the second-to-last equality follows from Table E.

4. (a) Note that many acecptable responses are possible. One such response is as follows.

Yes, sort of. The treatment difference caused different responses, but the actual "treatment" received in the vitamin C group was both a daily dose of Vitamin C and knowledge that it was vitamin C. It is possible that the second aspect of this treatment is what was responsible for the difference. Researchers must make sure that the two groups are treated as similarly as possible in all respects, except for the specific agent under comparison.

4. (b) Note that many acecptable responses are possible. One such response is as follows.

No, these results cannot be generalized to other 14-to-16 year olds since these are *not* randomly selected marijuana users, but instead are volunteers. In order for the results of this study to be applicable to (the population of) all 14-to-16 year olds, a random sample (and not a self-selected sample) must have been taken from the group of all 14-to-16 year olds. A number of different sampling methods could have been used to achieve this such as stratified sampling by either region or socio-economic status, etc.