Statistics for the Life Sciences

Math 20340 Section 01, Fall 2009

Homework 6 Solutions

- **7.19**: In each case mean of $\bar{x}$ is population mean, and standard deviation of $\bar{x}$ equals population std dev over square root of sample size.
  - **a**: Mean 10, std dev $.5$
  - **b**: Mean 5, std dev $.2$
  - **c**: Mean 120, std dev $1/\sqrt{8} = .35355...$
- **7.20**: See box of page 266, "How do I decide when the sample size is large enough"?
  - **a**: $\bar{x}$ will be *exactly* normal in all cases
  - **b**: $\bar{x}$ will be approximately normal for a) and b); but we can't say anything about c), since n is too small
- **7.25**:
  - **a**: Mean 106, std dev 2.4
  - **b**: $P(\bar{x} > 110) = P((\bar{x} - 106)/2.4 > (110 - 106)/2.4) = P(z > 1.667) = 1 - P(z \leq 1.667) = 1 - .9525 = .0475$
  - **c**: $P(102 \leq \bar{x} \leq 110) = P(-1.67 \leq z \leq 1.67) = .9525 - .0475 = .905$
- **7.26**:
  - **a**: $\bar{x}$ is approximately normally distributed, with mean (presumably) 64571 and standard deviation 4000
    $sqrt60 \approx 516.4$.
  - **b**: Within ± 1.96 standard errors of mean; so between 63558 and 65583.
  - **c**: 66000 is about 2.77 standard errors above mean, so the probability is about $.0028$ (very low).
  - **d**: Yes, given the answer to last part. Possible conclusions:
    - Random occurrence (a probability $.0028$ event does occur on average one every $1/.0028 = 357...$ times).
    - Sample perhaps not representative.

      * The way annual salary was reported in the data collection differed from the way it was report to the source that calculated the 64571 average.
      * The source citing an average of 64571 may be outdated and no longer valid.

- **7.27**:
  - **a**: Air temperature, time at which measurement is taken, variations in amounts of the various substances introduced initially, ...
  - **b**: Large number; variability of average error (measured by std dev) = variability of single measurement over square root of number of measurements; larger number of measurements leads to smaller variability.
- **7.29**: Think of a cubic foot of water as being made up of many (1728) cubic inches. Total number of bacteria is equal to sum of numbers in each square inch. It's reasonable to assume that the numbers in different square inches are independent, so total is sum of results of large number of independent experiments, each with same mean and std dev. Central limit theorem applies to say that the sum is approximately normal.
- **7.31**: Let x_i be amount of Potassium in banana i. Each of x_1, x_2 and x_3 are independent normal random variables with mean 630 and std dev 40.
  - **a**: $T = x_1 + x_2 + x_3$; by our sum variant of central limit theorem this has mean $3 * 630 = 1890$ and std dev $\sqrt{3} * 40 = 69.28$.
  - **b**: $P(T > 2000) = P(z > 1.59) = .0559$ (2000 is 1.59 standard deviations above 1890).
- **7.33**:
  - **a**: Distribution of sample mean for sample of size 130 has mean 98.6, std dev $.8/\sqrt{130} = .07....$ 98.25 is .35 below mean, or 5 standard deviations, so probability is very low (very close to zero).
  - **b**: Very unlikely.
- **7.37**:
  - **a**: mean .3, std dev .0458...
  - **b**: mean .1, std dev .015...
  - **c**: mean .6, std dev .03...
- **7.38**:
  - **a**: No, $np = 2.5 \leq 5$
  - **b**: Yes, $np = 7.5$, $nq = 67.5$, both > 5
  - **c**: No, $nq = 2.5 \leq 5$

- **7.41**: $SE(\hat{p}) = \sqrt{\hat{p}\hat{q}/n}$ in all cases.
  - **a**: .0099...
  - **b**: .03
  - **c**: .0458...
  - **d**: .05
  - **e**: .0458...
  - **f**: .03
  - **g**: .0099...
  - **h**: Max at $p = .5$; close to zero for p close to 0 or 1
- **7.45**:
  - **a**: Approximately normal with mean .13, standard deviation $\sqrt{(.13 \times .87)/55} = .0453$.
  - **b**: About .9382
  - **c**: Almost 0
  - **d**: Between .04 and .22
- **7.47**:
  - **a**: $p = .75$, $n = 200$, so mean of $\hat{p}$ is $p = .75$ and standard deviation is $\sqrt{pq/n} = .03...$; distribution is approximately normal
  - **b**: Greater than 80% is .05 above mean, or 1.67 standard deviations. Probability of this is .0475.
  - **c**: 95% of the time the sample proportion will be within ± 1.96 std devs of mean, so in range $.75 - 1.96 * .03$ to $.75 + 1.96 * .03$ or .6912 to .8088.