STA 580 — Spring 2011 — Dr. Charnigo

Written Assignment 1 Solutions

1a. Among exercising adults, the sample mean and sample standard deviation are 129.04 and 12.77, respectively. [Details for pencil-and-paper calculations: We have $n = 24$, $\sum_{i=1}^{n} x_i = 3097$, $\sum_{i=1}^{n} x_i^2 = 403391$, and $(\sum_{i=1}^{n} x_i)^2 = 3097^2 = 9591409$. Hence, we find that $\bar{x} = 3097/24 = 129.04$, $s^2 = (403391 - 9591409/24)/23 = 163.00$, and $s = \sqrt{163.00} = 12.77$.]

Among non-exercising adults, the sample mean and sample standard deviation are 144.80 and 14.93, respectively. [Details for pencil-and-paper calculations: We have $n = 25$, $\sum_{i=1}^{n} x_i = 3620$, $\sum_{i=1}^{n} x_i^2 = 529528$, and $(\sum_{i=1}^{n} x_i)^2 = 3620^2 = 13104400$. Hence, we find that $\bar{x} = 3620/25 = 144.80$, $s^2 = (529528 - 13104400/25)/24 = 223.00$, and $s = \sqrt{223.00} = 14.93$.]

1b. Among exercising adults, the sample median and sample interquartile range are 128.5 and 136.5-119=17.5, respectively. [Details for pencil-and-paper calculations: Since $n = 24$ is even, the median is the average of the 12th and 13th ordered observations; this is $(128 + 129)/2 = 128.5$. To find the 75^{th} percentile, note that $p = 75$ and that $np/100 = 18$ is an integer. The greatest integer less than or equal to 18 is $k = 18$, so the 75^{th} percentile is the average of the 18th and 19th ordered observations; this is $(135 + 138)/2 = 136.5$. To find the 25^{th} percentile, note that $p = 25$ and that $np/100 = 6$ is an integer. The greatest integer less than or equal to 6 is $k = 6$, so the 25^{th} percentile is the average of the 6th and 7th ordered observations; this is $(119+119)/2 = 119$. The interquartile range is the difference between the 75^{th} and 25^{th} percentiles; this is 17.5.]

Among non-exercising adults, the sample median and sample interquartile range are 145 and 150-134 = 16, respectively. [Details for pencil-and-paper calculations: Since $n = 25$ is odd, the median is the 13th ordered observation; this is 145. To find the 75^{th} percentile, note that $p = 75$ and that $np/100 = 18.75$ is not an integer. The greatest integer less than or equal to 18.75 is $k = 18$, so the 75^{th} percentile is the 19th ordered observation; this is 150. To find the 25^{th} percentile, note that $p = 25$ and that $np/100 = 6.25$ is not an integer. The greatest integer less than or equal to 6.25 is $k = 6$, so the 25^{th} percentile is the 7th ordered observation; this is 134. The interquartile range is the difference between the 75^{th} and 25^{th} percentiles; this is 16.]

1c. [You should display a printout of the side-by-side box plots.] The sample distribution of systolic blood pressure among non-exercising adults is modestly but not markedly asymmetric. The distance from the 25^{th} percentile to the 50^{th} percentile is somewhat greater than the distance from the 50^{th} percentile to the 75^{th} percentile, which would seem to suggest a left skewed distribution. However, the lower "whisker" is of similar length to the upper "whisker".

The sample distribution of systolic blood pressure among exercising adults is very close to symmetric. The lower "whisker" is of similar length to the upper "whisker", and the distance from the 25^{th} percentile to the 50^{th} percentile is similar to the distance from the 50^{th} percentile to the 75^{th} percentile.

Regarding central tendency, there is moderate distinction between the two sample distributions: larger values of systolic blood pressure occur more often among non-exercising adults. Regarding variability, there is little distinction between the two sample distributions.

1d. The population percentage of exercising adults who have systolic blood pressure measurements between 120 and 140 is $P(120 \leq X \leq 140)$, where X is normal with mean 129.04 and standard deviation 12.77. By standardization, the requested probability is $P\left(\frac{120-129.04}{12.77} \leq Z \leq \frac{140-129.04}{12.77}\right)$, where Z is standard normal. Using Table 3 or SAS, we find that $\Phi\left(\frac{140-129.04}{12.77}\right)-\Phi\left(\frac{120-129.04}{12.77}\right)=\Phi(0.858)-\Phi(-0.708)=0.805-0.239=0.566=56.6\%$.

The population percentage of exercising adults who have systolic blood pressure measurements above 140 is $P(X>140)=1-P(X\leq 140)=1-P\left(Z\leq \frac{140-129.04}{12.77}\right)=1-\Phi(0.858)=0.195=19.5\%$.

1e. The systolic blood pressure measurement defining the boundary between the top third and the bottom two-thirds in the population of exercising adults is the number c such that $P(X\leq c)=2/3$, where X is normal with mean 129.04 and standard deviation 12.77. By standardization, we have $P\left(Z\leq \frac{c-129.04}{12.77}\right)=\Phi\left(\frac{c-129.04}{12.77}\right)=2/3$. On the other hand, Table 3 or SAS shows that $\Phi(0.431)=2/3$, so $0.431=(c-129.04)/12.77$. Solving for c yields 134.54.

2a. Let A denote the event that an adult is obese, and let B denote the event that the adult is diabetic. We are told that $P(A)=0.30$, $P(B|A)=0.15$, and $P(B|\bar{A})=0.075$. Note immediately that we must have $P(\bar{A})=1-0.30=0.70$, $P(\bar{B}|A)=1-0.15=0.85$, and $P(\bar{B}|\bar{A})=1-0.075=0.925$.

Our first task is to find $P(B)$. We apply the law of total probability,

$$P(B)=P(B|A)P(A)+P(B|\bar{A})P(\bar{A})=0.15\times 0.30+0.075\times 0.70=0.045+0.0525=0.0975.$$

Thus, 9.75% of adults are diabetic.

Our second task is to find $P(\bar{B})$. This equals $1-P(B)=1-0.0975=0.9025$. So, 90.25% of adults are not diabetic.

2b. Our first task is to find $P(A\cap B)$. This equals $P(B|A)P(A)=0.15\times 0.30=0.045$. Thus, 4.5% of adults are both obese and diabetic.

Our second task is to find $P(\bar{A}\cap\bar{B})$. This equals $P(\bar{B}|\bar{A})P(\bar{A})=0.925\times 0.70=0.6475$. So, 64.75% of adults are neither obese nor diabetic.

2c. Our first task is to find $P(A|B)$. We apply Bayes' Theorem,

$$P(A|B)=\frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|\bar{A})P(\bar{A})}=\frac{0.045}{0.0975}=0.462.$$

Among diabetic adults, 46.2% are obese.

Our second task is to find $P(A|\bar{B})$. We apply Bayes' Theorem,

$$P(A|\bar{B})=\frac{P(\bar{B}|A)P(A)}{P(\bar{B}|A)P(A)+P(\bar{B}|\bar{A})P(\bar{A})}=\frac{0.85\times 0.30}{0.85\times 0.30+0.925\times 0.70}=0.283.$$

Among non-diabetic adults, 28.3% are obese.

2d. Let X be the number of adults, among 20 randomly selected, who have diabetes. Then X is a binomial random variable with $n=20$ and $p=0.0975$. Our task is to find $P(X\geq 3)$. Noting that this equals $1-P(X=0)-P(X=1)-P(X=2)$, we can employ SAS or the formula for the probability mass function of a binomial distribution. Either way, we find that $P(X=0)=0.1285$, $P(X=1)=0.2777$, and $P(X=2)=0.2850$. From this we conclude that $P(X\geq 3)=1-0.1285-0.2777-0.2850=0.3088$.

Let Y be the number of adults, among 200 randomly selected, who have diabetes. Then Y is a binomial random variable with $n=200$ and $p=0.0975$. Our task is to approximate $P(Y\geq 30)=P(30\leq Y\leq 200)$. Since $np(1-p)=17.6\geq 10$, we are comfortable invoking the Central Limit Theorem to obtain the approximate probability. Noting that $\Phi(x) \approx 1$ for $x \geq 4$, we obtain

$$\begin{aligned} P(30 \leq Y \leq 200) &\approx \Phi\left(\frac{200.5/200 - 0.0975}{\sqrt{0.0975(0.9025)/200}}\right) - \Phi\left(\frac{29.5/200 - 0.0975}{\sqrt{0.0975(0.9025)/200}}\right) \\ &= \Phi(43.15) - \Phi(2.384) \\ &\approx 1 - 0.9914 \\ &= 0.0086. \end{aligned}$$

2e. Let Y denote the number among 5000 randomly selected adults who have diabetes. Then Y is a binomial random variable with $n = 5000$ and $p = 0.0975$. The expected value of Y is $5000 \times 0.0975 = 487.5$, the variance of Y is $5000 \times 0.0975 \times 0.9025 = 440.0$, and the standard deviation of Y is $\sqrt{440.0} = 20.98$.

So, if 200 out of 5000 adults in a community have diabetes, that is less than the 487.5 we would expect. Actually, the 200 is more than 13 standard deviations less than the 487.5 we would expect. Hence, this finding is quite surprising. [The probability that a normal random variable is more than 13 standard deviations away from its mean is essentially zero. The same is true for a binomial random variable when n is large enough for us to invoke the Central Limit Theorem.]