Solutions to Exam #2

Math 114-D

Friday, April 9, 1999

1. Sampling bias occurs when your method of choosing a sample favors some members of the population over others. One possible example would be sampling habits of automobile drivers by surveying owners of Dodge minivans.

2. The regression model is $y = 21.8627 - 0.1030x$. The proportion of variability explained by the model is $r^2 = 0.9268 = 93\%$. The equation for the regression line explains 93% of the variability of the data y-values; the model is a very good fit.

3. The regression model is $y = -1139.5243 + 92.5146x$. The fitted values are the y's gotten by plugging the data set's x's into the regression model.

x	$-1139.5243 + 92.5146x$
11	-121.8641
15	248.1942
18	525.7379
14	155.6796
19	618.2524

4. If $\Pr(X > k) = 0.15$, then $\Pr(X \leq k) = 0.85$.

   According to Table I, the z with 0.85 below it is 1.04.

   The raw score corresponding to $z = 1.04$ is $55 + (1.04)(12) = 67.48$.

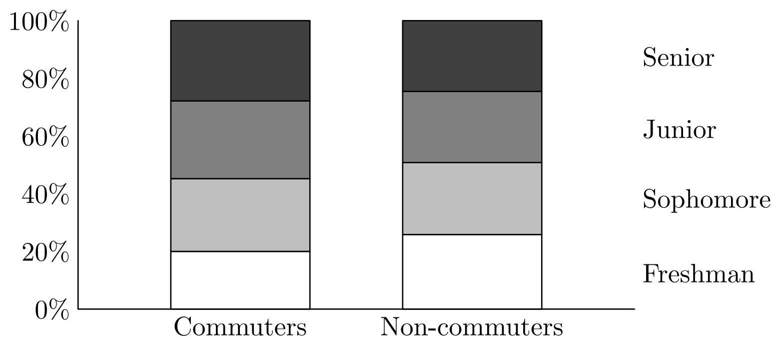
   $\Pr(X > 67.48) = 0.15$.

5. There are 575 commuters. The conditional distribution for commuters is found by dividing by 575:

	Freshman	Sophomore	Junior	Senior
Commuter	20.00%	25.22%	26.96%	27.83%

   There are 3425 non-commuters. The conditional distribution for non-commuters is found by dividing by 3425:

	Freshman	Sophomore	Junior	Senior
Non-commuter	25.84%	24.96%	24.67%	24.53%



6. A residual is the difference between the actual y value and the one predicted by the regression model for a given data point (used in the construction of the regression model).

   If there is a pattern visible in the residual plot, then there is some factor at work that is not accounted for by our model. Such a pattern means that the model needs to be revised.

7. The standard deviation of $\hat{p}$ is given by $\sqrt{\frac{\theta(1-\theta)}{n}}$. It decreases as n grows larger, and follows a curve from 0 at $\theta = 0$ up to a peak at $\theta = 0.5$ and back down to 0 at $\theta = 1$.

8. $\hat{p}$ is normal with mean 0.58 and standard deviation $\sqrt{\frac{0.58(1-0.58)}{325}} = 0.0274$.

$$
\begin{aligned}
\Pr(\hat{p} > 0.6) &= \Pr\left(z > \frac{0.6 - 0.58}{0.0274}\right) \\
&= \Pr(z > 0.73) \\
&= 1 - \Pr(z \leq 0.73) \\
&= 1 - 0.7673 \\
&= 0.2327 \\
&= 23.3\%
\end{aligned}
$$

9. (a) Samples are smaller, so it is cheaper and faster.

   (b) Sometimes measuring something destroys the thing measured; it is better to destroy the items in a sample than to destroy the whole population.

10. The Central Limit Theorem states that any variable that involves summing or averaging the same thing repeatedly will have a normal distribution. The Theorem works because high and low values tend to cancel each other out.