## **General Problems for Review**

Starting on page 265

#13 (i) it's nearly zero. (ii) it's nearly 1 (iii) nearly -1

#15 About 68% of the data is within plus or minus 10 points of the regression line and 16% is above the line by 10 points or more, the other 16% is below by more than 10 points.

#17 The 25<sup>th</sup> percentile is about .67 in standard units (.65 or .70 is close enough from your table A 105). This would be about a 73 in the original units. In other words, of the students who scored around 50 on the midterm, you need to find the percentage scoring over 73 on the final. The new average would be about 60.

#18 There are 30 possible outcomes.

(a) There are two good outcomes (2 5) and (5 2) their chance is 2/30

(b) 5/30

(c) 10/30

Starting on Page 428

#1 No. Teachers might be tempted to put poorer children into the treatment group.

#3 (a) Yes. Other things being equal, drinkers might have higher rates of oral cancer than nondrinkers because of alcohol

(b) ii

#8 Yes.

#9 Guess about 520. You would have a 68% chance to be within about 86 points.

#11 (a) .30 (b) probably too high. You need to use the other line, for predicting husband from wife.

#16 False. Option (i) has chance  $1-(37/38)^{15}$  about 33% while option (ii) is like 55%

#20 (a) using the normal will give a result which is too low because the normal curve is lower than the histogram at 90. (b) using the normal is about right because the highs and lows will cancel each other out.

#27 (a) about 10.8% (b) about 1.4% (c) the range of 8% to 13.6% is a 95% confidence interval for the percentage of independents among all registered voters in Hayward.

#30 False -- accuracy in an estimate percentage depends on the sample size, not on the population size. You can tell this by looking at the formula for the standard error of the percentage. What determines its size is primarily the number of draws.

Starting on Page 565

#3 False. You need to take the same size into account too. For the first investigator, you will get a Z around 2 and therefore a p-value around 5%. The second one will get a Z around 3 and a p-value around 0.3%

#4 Yes. Simply by chance you could get a statistically significant z-test without doing anything except drawing random samples.

Starting on Page 567

#1 (a) observational study (b) treatment is prisoners who finished boot camp. Control are drop outs. (c) those who stayed might be different to begin with than those who eventually drop out.(d) (i) the treatment group consists of all those who volunteer for the program whether or not they complete it. The control group consists of all non-volunteers. (ii) the recidivism rate is similar suggesting there is no effect of the program. (e) Most completed.

#2 Not good. The comparison is between right and left handers. A confounder must be correlated with left handedness among players and cause mortality.

#8 y = .533x + 1.667

#9 (a) A will get a higher correlation, B suffers from a restriction in range, (b) the correlation for averages will be higher

#11 A first-year GPA of 3.5 is 1SD above average. Students with this GPA averaged about  $r^*1 = 0.4$  SDs higher in the second via the regression method. Sally must have been above average by about .4 SDs putting her in the 66<sup>th</sup> percentile.

#15 (a) 4/52 \* 3/51 \* 2/50 = 2/10000 (b) 48/52 \* 47/51 \* 46/50 = 78% © 36/52 \* 35/51 \* 34/50 = 32% (d) 100% - 32% = 68%

#17 The net gain is like the sum of 100 draws from a box with 18 tickets marked +1 and 18 marked -1 and 2 tickets marked -.50. The average of this box is -1/38 or .0263. The expected net gain is -2.63. The SD of the box is .98. The SE for the net gain is 9.80. So 2.63/9.80 is about .27 or the answer is about 40%

#18 Using telephones tends to exclude the homeless, the recently homeless and the very poor. On this basis, 3% might be too low.

#20 The number of families without cars is like the sum of 1500 draws from a 0,1 box. There is a ticket in the box for each of the 25,000 families in the town marked 1 (no car) 0 (owns cars). SO the fraction of 1's in the box is .1 and the SD of the box root(0.1 \* 0.9) = 0.3 so the expected value for the sum is 1500 \* 0.1 or 150. The SE is root(1500) \* 0.3 = 12. The number of sample families without cars will be around 150 give or take 12 or so. 12 out of 1500 is 0.8% So the percentage of sample families without cars is 10% give our take 0.8% or so. The chance will be about 80%

#22 The town is large there is not much difference between drawing with and without replacement. The number of phones is like the sum of 500 draws form a box with 20 tickets market 1 and 80 marked 0. The expected number will be 100. The SE is 8.94 and so 0.5/8.94 is about 0.056. If you look up table A, you will get about 4%.

#26 (a) True. (b) True. (c) False. (d) True see page 411, 418-419. (e) False (f) False