Fall Term 2014
STAT 2507 D
Solutions to Assignment # 2

Part I: Lab Questions

1. **[20]** To verify the widely held belief that sales price is related to taxes paid, the data on the selling price and annual taxes for 24 houses have been collected (see the Excel file attached). Copy and paste the data in this column in a Minitab spreadsheet (which you need to open first).

   a. Construct a scatterplot with *price* marked along the horizontal axis and *taxes* marked along the vertical axis. Does the scatterplot present a linear relationship between the variables *price* and *taxes*? **[5]** Yes, the points on the scatterplot band around a line with a positive slope.

   b. Calculate the correlation coefficient r **[2]**: $r = 0.876.$ Does the value of r indicate a strong linear relationship between *price* and *taxes*? Explain. **[3]** Yes. The value 0.876 is close to one. This, together with our visual impression of the distribution of the data obtained from the scatterplot, indicates a strong positive linear relationship between the two variables.

   c. If appropriate, fit a least squares (LS) regression line using *price* (the predictor variable) to predict the *taxes* (the response variable). What is the equation of the LS regression line? **[5]** $\textit{Taxes} = -1.584 + 0.2308\,\textit{price}$. **Comment:** The points on the scatterplot band around a line. The correlation coefficient confirms a strong positive linear relation between the two variables. Hence *taxes* can be predicted from *price* by fitting a LS line.

   d. What is the predicted *taxes* corresponding to the *price* 32.4? **[1]** 5.8939. Corresponding to the *price* 36.1? **[1]** 6.7479. Corresponding to the *price* 49.8? **[1]** 9.9098. Which of these seem to make sense? **[2]** For the values in the range of the data, 32.4 and 36.1.

2. The article "Human Lateralization from Head to Foot: Sex-related Factors" (*Science*, 1978: 1291–1292) reports for both a sample of right-handed men and a sample of right-handed women the number of individuals whose feet were the same size, had a bigger left than right foot (a difference of half a shoe size or more), or had a bigger right than left foot.

   [9] (1) Exact probabilities. Suppose that one of the 127 individuals is selected at random. What is the probability that the individual

   a. is a man? $P(M)$ =40/127=0.315.

   b. has a bigger left than right foot? $P(A)$ =57/127=0.449.

   c. is a man and has a bigger left than right foot? $P(M \cap A)$ =2/127=0.016.

   (2) Probabilities as relative frequencies. In this question, we see how relative frequencies can approximate the true probabilities calculated in the previous questions. To do so, we repeat the process of selecting a single person many times, using **sampling with replacement**; that is, each time we select a single individual from the **same** population of 127 individuals. We want to investigate the accuracy of the relative frequency estimates in two cases: when the number of times we repeat the sampling process is small (say, between 15 and 30) and when the number is large. We use MINITAB to assist us in doing so.

NOTE: The estimated probabilities in Lab part may vary from student to student. To give the estimated probabilities below, I use the results from my MINITAB session (the simulation results are attached).

[9] **I.** Use MINITAB to simulate 20 replications of the experiment that consists of selecting a single person from the population of interest. For the 20 replications you simulated, what are the relative frequencies of individuals that

a. are men? $\hat{P}(M) =$ (1+4)/20=0.250.

b. have a bigger left than right foot? $\hat{P}(A) =$ 7/20=0.350.

c. are men and have a bigger left than right foot? $\hat{P}(M \cap A) =$ 0.

(**Note 2:** The symbol "ˆ" denotes that the relative frequency is **only** an estimate of the **actual** probability. $\hat{P}$ reads: P hat.)

[9] **II.** Repeat part I, using 400 replications.

a. $\hat{P}(M) =$ (7+34+90)/400=0.328.

b. $\hat{P}(A) =$ (7+175)/400=0.455.

c. $\hat{P}(M \cap A) =$ 7/400=0.018.

[6] **III.** Compare the relative frequencies you obtained in parts I and II above with the actual probabilities you obtained in question 1. Are they close? Comparing the actual and estimated probabilities, we observe that for a large number or replications the relative frequencies estimate the actual probabilities very well, whereas in case of the a small number of replications, the approximation is worse. This is consistent with the definition of probability as the **limiting** relative frequency. In general, the larger the sample size is, the better the approximation is.

[6] **(3) Conditional probabilities.** If an individual is selected at random from the 127 individuals under study, find

a. the probability that the individual is a man given that feet are the same size; $P(M|B) =$10/28=0.357.

b. the probability that feet are the same size given that the individual is a man; $P(B|M) =$10/40=0.250.

[9] **(4) Relative frequencies.** To estimate the probability in question 3a, we reduce the sample space by removing all of the individuals whose feet are not the same size. Our population now contains 28 individuals (corresponding to the second column of the contingency table).

a. Use MINITAB to simulate 20 replications of the experiment that consists of selecting a single person from the reduced population and compute the relative frequency of the individuals whose feet are the same size and who are men. $\hat{P}(M|B) =$ 5/20=0.250.

b. Use MINITAB to simulate 400 replications of the experiment that consists of selecting a single person from the reduced population and compute the relative frequency of the individuals whose feet are the same size and who are men. $\hat{P}(M|B) =$152/400=0.380.

c. Is the relative frequency using larger number of replications closer to the actual value $P(M|B)$ found in question 3a? Would you expect it to be so? Explain. The estimated values of $\mathbf{P}(M|B)$ for a large number of replications is much closer to the theoretical value of $\mathbf{P}(M|B)$

than for a small number of replications. In general, a large-sample approximation is expected to be better than a small-sample approximation.

[9] **(5) Relative frequencies.** Now, in order to estimate $P(B|M)$, we reduce the sample space by removing all of the women from the population under study. The reduced population contains 40 men (corresponding to the first row of the contingency table).

a. Use MINITAB to simulate 20 replications of the experiment that consists of selecting a single person from the reduced population and compute the relative frequency of the individuals whose feet are the same size and who are men. $\hat{P}(B|M) =$ 3/20=0.150.

b. Use MINITAB to simulate 400 replications of the experiment that consists of selecting a single person from the reduced population and compute the relative frequency of the individuals whose feet are the same size and who are men. $\hat{P}(B|M) =$ 100/400=0.250.

c. Is the relative frequency using larger number of replications closer to the actual value $P(B|M)$ found in question 3b? Would you expect it to be so? Explain. Yes, the estimated value of $\mathbf{P}(B|M)$ for the large number of replications is much close to the theoretical value of $\mathbf{P}(B|M)$. This confirms a general fact that a large-sample approximation is typically better than a small-sample approximation.

Part II: Non-lab questions

3. The outcome b means that a child is a boy, and g means that it is a girl.

   a. The sample space is $S = \{b, g, bg, gb, bb, gg, bbb, bgb, bbg, gbb, bgg, gbg, ggb, ggg\}$.

   b. We have $A = \{g, gb, gg, ggg, ggb, gbg, gbb\}$ and $B = \{gg, bgg, gbg, ggb\}$. Since $A \cap B = \{gg, ggb, gbg\} \neq \emptyset$, it follows that A and B are not mutually exclusive.

4. We have

   a. $\mathbf{P}(A \cup B) = P(A) + P(B) - P(A)P(B) = 0.20 + 0.62 - 0.124 = 0.696.$

   b. $\mathbf{P}(A \cup B) = P(A) + P(B) = 0.20 + 0.62 = 0.82.$

   c. We have $A = A \cap S = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c) = \emptyset \cup (A \cap B^c) = A \cap B^c$, and hence $P(A \cap B^c) = P(A)$. From this, $P(A|B^c) = \dfrac{P(A \cap B^c)}{P(B^c)} = \dfrac{0.20}{1 - 0.62} = 0.526.$

5. **a.** The desired probability is $\dfrac{C_2^4}{C_2^{10}} = \dfrac{6}{45} = 0.133.$

   b. The desired probability is $(4/10)(4/10) = 16/100 = 0.16.$