1. Health insurance claim processing

(a) Either a yes or no answer are both acceptable with a good explanation. The comment that the variable is integer suggests that a normal is not likely to be a good approximation because the normal is used for continuous quantities. You might also note that the normal distribution takes values in \((-\infty, \infty)\) while the number of days is positive. That being said, the number of values that the random variable can take is quite large (1,2,\ldots,100,101,\ldots) and the mean is substantially larger than the standard deviation which suggests a normal approximation might be adequate.

(b) Let \(X\) denote the number of days required to process a claim. Assume \(X \sim N(32,8)\). Then \(\Pr(X > 40) = \Pr(Z > 1)\) where \(Z\) is a standard normal. This is equal to 0.16. Similarly, \(\Pr(X > 50) = .012\) and \(\Pr(X > 60) = .00023\). Some people noted that because \(X\) is discrete it would be better to use the continuity correction and compute \(\Pr(X > 39.5), \Pr(X > 49.5), \Pr(X > 59.5)\).

(c) My thinking here is that the population of interest are claims processed with the new software and the population parameter of interest is the mean processing time.

(d) Under the hypothesis of no change, the claim processing times still have a mean of 32 days and a standard deviation of 8 days. The mean of 50 samples from such a population has mean 32 and s.d. \(8/\sqrt{50} = 1.13\) and should follow a normal distribution. We know this because of the Central Limit Theorem. (The question is a bit ambiguous because if you are still assuming a normal distribution as in part (b) - this was not my intent - then you know it has a normal distribution because of properties of the normal distribution.

(e) Here we want to know \(\Pr(X < 30.4)\) under the assumption of no change in processing times. We can use the result of the previous part to find that \(\Pr(X < 30.4) = \Pr(Z < (30.4 - 32.0)/1.13) = .079\). This probability is small which suggests there is some benefit to the new software; note that it is small but not very small so there is still a chance that the mean processing time hasn’t changed and we just got a lucky sample.

(f) From (b) we know that we only expect 1.2% of claims to take more than 50 days but had two out of 50 (4%) in our sample. If we think statistically here, we might assess how unusual this is by asking how likely it would be to get two (or more) successes if we were flipping a coin with probability of success 1.2%. According to the binomial distribution there is a 12% chance of seeing two long claim processing times out of 50 so we should probably not be overly concerned.

2. Experiment or observational study

(a) The experimental units are the students in the school. The treatments are milk or no milk (two groups). The response variable for each unit is the reading level at the end of the year.

(b) This is an experiment. The investigators assigned the children to the different treatment groups.

(c) Random assignment guarantees that the two groups of children (milk, no milk) have approximately the same distribution of characteristics in all respects (i.e., height, weight, gender) except for the treatment received. Thus any difference we see in the mean response is either due to chance or the treatment.

i. The study started out as an experiment. When the teachers changed treatment assignments it became more of an observational study.

ii. ITT - The advantage here is that this corresponds to the randomized experiment. These two groups of students should match on all other features (race, gender, family income, etc.). The disadvantage is that because some of the assigned treatment group did not actually get their milk and some of the assigned control group did get milk, we are likely to see a smaller impact than we might have otherwise.

iii. As treated - The advantage here is we are actually measuring the effect of receiving milk rather than the effect of "being assigned to receive milk" as in the ITT analysis. The big disadvantage is that the two treatment groups are not comparable because of the teachers’ interference; the milk group now includes more needy students than the no milk group. Analyzing these groups is like analyzing an observational study. If we see a difference in outcomes we don’t know whether it is due to the treatment or to the differences between the groups. In fact, by giving milk to the needy students in the control group and taking milk away from the not-needy students in the treatment group the teacher may have made milk seem less effective!

iv. Per protocol - Per protocol analysis is sometimes thought of as a “best case” analysis because it compares the treatments using only those individuals that actually received their assigned treatment. It thus ignores people who don’t comply with their assigned treatment for whatever reason (here it is because of the teachers). Unfortunately these two groups still differ in ways that we don’t know because they are not the original randomized samples; thus it’s essentially an observational study.
3. P-values:

(a) It is very important to me that you learn how to carefully describe the $p$-value. A $p$-value of .04 means that the probability of observing a result as or more extreme than the one we obtained, if the null hypothesis is true, is .04. That is the end of the statement. It is incorrect to say that because the $p$-value is less than .05 the result is significant; it is incorrect because nobody told you to use .05 as a threshold and nobody asked whether the result was statistically significant.

   i. I’d be content to get rid of the term statistical significance. I share the concern expressed in these articles that science has become too dependent on the term and over interprets both statistical significance and its absence. That being said there is another side to the argument. If we don’t think about thresholds/decisions then anyone can carry out a statistical test and claim the result is important.

   ii. P-values measure what they measure; they assess the chance of seeing data like the observed data if the null hypothesis is true. They thus provide some insight into whether the data are consistent with that hypothesis. This is a positive aspect of $p$-values. Many of the negative features associated with $p$-values arise because of the way people interpret findings. By focusing on the binary decision (significant or not) they tend to either reject or accept the null hypothesis. That decision ought to depend on many other things beside the $p$-value. Among other things it is very important to remember that: (a) a significant result may not be practically important; and (b) a non-significant result doesn’t mean the null hypothesis is true.

4. Installing R and RStudio - Great work!