1. Normal distribution - This was intended as a review of the normal distribution.
   (a) If the weight $Y$ is $N(\mu = 45, \sigma = 7)$, then $P(40 < Y < 60) = P(-.714 < Z < 2.143)$ where $Z$ is a standard normal. From $R$ using the `pnorm` function $P(40 < Y < 60) = P(Z < 2.143) - P(Z < -.714) = .984 - .237 = .747$.
   (b) To find $y$ such that $P(Y < y) = .85$, we note that this would be equivalent to finding $y$ such that $P(Z < (y - 45)/7) = .85$. Using the `qnorm` function in $R$ we find that $(y - 45)/7 = 1.036$ and $y = 52.255$.

2. Light bulb quality
   (a) It is important to remember that our various statistical inference procedures (confidence intervals, point estimates, significance tests) all relate to some population quantity of interest (also known as a parameter). Here the company is interested in the defective rate of bulbs. It is most precise to say that we are interested in the parameter $\mu$ which is equal to the mean number of defective light bulbs per box in the population of manufactured boxes.
   (b) The normal distribution is a continuous distribution on the real line. The number of defectives in a box is a discrete random variable. Note also that the normal distribution would assign some probability to negative numbers and that doesn’t make any sense here.
   (c) According to the central limit theorem $M$ is approximately normal with mean equal to 1.4 and standard deviation $1.2/\sqrt{36} = 0.2$.
   (d) $P(M > 1.7) = (Z > 1.5) = .067$. Either decision (shutting down production or not shutting down production) can be justified by this result. The observed value (1.7) is unusually large but not necessarily too large to have occurred by chance. To actually decide whether to stop production we’d want to know the cost to the company of shutting down production (perhaps delaying delivery) and the cost to the company of shipping poor quality goods. Estimates of those costs might allow an informed decision about how to choose a cutoff for the $p$-value.

3. 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 neither an experiment or an observational study. It started out as an experiment. When the teachers changed treatment assignments it became more of an observational study. If I had to call it one thing, I would call it an observational study.
   (c) Random assignment guarantees that the two groups (milk, no milk) are the same in all respects (on average) except for the treatment received. Thus any difference we see is either due to chance or the treatment.
   (d) 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.
   (e) 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 an 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!
   (f) 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.

4. P-values: My personal view is that there is no reason to ban p-values. 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. Many of the problems 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 has 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.