18.650 Problem set 1 due Wednesday, September 16, 2015,
based on Chapter 6 and Section 8.5.3 of Rice, and class handouts:
"Normal distributions and sample statistics," normalsamples.pdf, and
"Confidence intervals for normal parameters," normalconfints.pdf

1. (a) Rice, p. 198, Problem 3, but with 36 instead of 16 and 0.6 instead of 0.5. More explicitly:

Let $\overline{X}$ be the average of a sample of 36 independent normal random variables with mean 0 and variance 1. Determine c such that $P(|\overline{X}| < c) = 0.6$.

(b) Rice, p. 198, Problem 4, but for t_5 instead of t_7. More explicitly:

If T follows a t_5 distribution, (i) find t_0 such that $P(|T| < t_0) = .9$, and (ii) find t_0' such that $P(T > t_0') = 0.05$.

The remaining problems will concern the Newtonian gravitational constant G. A Web search gave articles confirming that G is indeed a constant within measurement error. Suppose that $G = G(t)$ were a function of time. Investigators seem to assume that $d \log G(t)/dt = G'(t)/G(t)$ is a constant γ which might in principle be different from 0. An article "The confrontation between General Relativity and Experiment" by Clifford M. Will, in "Living Reviews in Relativity," lists values and ranges for γ from different methods, all *times* 10^{-13} *per year.*

Estimates of possible trend in G

Method	est. of γ and error bar
Lunar laser ranging	4 ± 9
Binary pulsar 1913+16	40 ± 50
Helioseismology	0 ± 16
Big bang nucleosynthesis	0 ± 4

Note that all the given ranges contain 0, with two centered at 0. As far as (astro)physicists have so far found, then, *the gravitational constant is indeed a constant.* It has proved surprisingly hard, however, as compared to other physical constants, to determine its value to many decimal places of accuracy.

2. The following measurements of G were published since 1998, in units of $10^{-11}\mathrm{N}\cdot\mathrm{m}^2/\mathrm{kg}^2$ (N=Newtons, m=meters, kg=kilograms).

est. $\pm$ error est.	year	Country and authors
6.6749 ± 0.0015	1999	Germany, Nolting et al.
6.6873 ± 0.0094	1998	USA, Schwarz et al.
6.6699 ± 0.0007	1999	China, Luo et al.
6.6742 ± 0.0006	1999	New Zealand, Fitzgerald and Armstrong
6.6830 ± 0.0011	1999	England, Richman et al.
6.6754 ± 0.0015	1999	Switzerland, Nolting et al.
6.674215 ± 0.000093	2000	USA, Gundlach and Merkowitz
6.67234 ± 0.00014	2010	USA, Faller and Parks

The last two measurements have much narrower error bars than the preceding ones, although they differ in their estimates of G by much more than their given error estimates. In this problem, ignore the error bars such as ± 0.0029 given by the experimenters and just consider the numbers in the leftmost column as data points. That leaves us with 8 estimates of G.
(a) Find the sample mean $\overline{X}$ of the 8 observations.
(b) Find their sample variance $S^2 \equiv s_X^2$ and standard deviation s_X. *Hint*: scientific calculators often give $s = s_X$; square it to get s_X^2. Or in R, for a data vector x, var(x) gives its variance and sd(x) its standard deviation.
(c) Find a 95% confidence interval for the gravitational constant G (the true mean of the observations), based on the 8 data points assuming a normal distribution.

3. For this problem, we'll study the errors, in other words the variances. For the 8 observations (still the numbers preceding "$\pm$"), assuming that they are i.i.d. normal with the same but unknown mean μ and variance σ^2,
(a) Give a 95% confidence interval for σ^2. Take the square roots of the endpoints to get a 95% confidence interval for σ.
(b) In this part, consider the estimated experimental errors for the eight observations, given after $\pm$. Which of them are within the confidence interval for σ? Is the 2010 CODATA value for "standard uncertainty" of G, .00080, in the confidence interval?

(c) As the last two experiments from 2000 and 2010 had substantially smaller error estimates, let's consider just the set of $n = 2$ estimates of G from those studies. Repeat parts (a) and (b) just for this smaller data set.

4. When an unknown constant, such as G, is estimated by a certain method or procedure, suppose we get a set of i.i.d. estimates $T_1, ..., T_n$. There may be a "bias" $b \neq 0$ in the method, so that the value $\mu = ET_j$ for all j is not G but $G + b$. The T_j will also have a variance σ^2. The given error bar amount after $\pm$ can be viewed as trying to estimate σ, or sometimes 2σ. Now consider the following three estimates of G from the 1890's:

6.658 ± 0.007,	1895	England, C. V. Boys
6.657 ± 0.013,	1896	Hungary, R. Eötvös
6.658 ± 0.002,	1897	Austria, C. A. Brayn

Assume that these estimates were all made by the same method.

(a) In light of the estimates in the table "Estimates of possible trend in G," is it plausible that the differences in the above three estimates from those in 1998 and afterward resulted from an actual change in G in a little over 100 years? *Hint*: Solve the differential equation $G'(t) = \gamma G(t)$ and from the given bounds for the constant γ, find by what factor G could have changed in the given amount of time.

(b) What was the approximate bias of the method used, relative to the "CODATA" value of G (2006, 2010) of 6.674 to the same number of decimal places? (By the way, the 1998–2010 estimates were made by several methods. They could have had biases, but much smaller ones than the 1895–1897 estimates had.)

(c) Which of the three experiments had the smallest error according to the declared values in each paper?

(d) With hindsight, which of the three had the most realistic error bar, in the sense of accounting for some possible bias as well as the σ within the given method?

(e) Do the results of the three studies appear to have been, in fact, independent? Why or why not?

5. (i) In each of the following cases, what can one conclude about the hypothesis the the data set is i.i.d. normal? Options are: RBT (reject by the Shapiro-Wilk test if the p-value is less than 0.05), RFO (reject for another reason – what reason?) A (accept the hypothesis), AP (accept provisionally, for purposes of analyzing the data).

(a) The data $x_1, ..., x_{25}$ are i.i.d. uniform $U[0, 1]$, as generated by x = runif(25) in R, and shapiro.test(x) gives a p-value 0.1964.
(b) $x_1, ..., x_{15}$ are i.i.d. standard exponential, having density e^{-x} for $x > 0$ and 0 otherwise, as generated by x = rexp(15) in R, and shapiro.test(x) gives a p-value 0.00272.
(c) $z_1, ..., z_8$ are observed real data, and shapiro.test(z) gives a p-value 0.1085.

(ii) To get insight into the different outcomes in parts (a) and (b), we can get a normal distribution with any real mean μ and variance σ^2, for a random variable X having either of the given distributions, or a $N(0, 1)$ distribution, evaluate the "skewness" $E(X - \mu)^3/\sigma^3$. In this regard, which of the distributions in (a) or (b) is more like a normal distribution?