# QBS 120 - Problem Set 4

Rob Frost

1. (Based on Rice 6.3) Let $\bar{X}$ be the average of a sample of $n$ independent standard normal RVs.
   (a) Determine $c$ such that $P(|\bar{X}| < c) = 0.5$. Solve for $c$ as a function of $n$.
   (b) Using only R *norm() functions for the standard normal distribution, compute the exact value of $c$ for $n = 5, ...100$ and visualize as a plot of $c$ vs. $n$.
   (c) If the variance was not known, how would you solve the problem and what additional piece of information would you need to get an exact answer?
   (d) If the $n$ RVs are independent and have the same distribution with expectation 0 and variance 1 but the exact distribution is not known, how would you approach the problem?

2. (Based on Rice 6.6)
   (a) Show that if $T \sim t_n$, then $T^2 \sim F_{1,n}$.
   (b) For n=10, demonstrate this equivalence numerically by plotting the kernel density estimates for 1000 randomly generated $T^2$ values and 1000 randomly generated $F_{1,n}$ values.

3. (Optional; based on Rice 6.9)
   (a) Find the mean of $S^2$, where $S^2$ is as defined in Section 6.3.
   (b) Find the variance of $S^2$, where $S^2$ is as defined in Section 6.3.

4. (Based on Rice 7.3) Which of the following is a random variable? Justify your answers.
   (a) The population mean.
   (b) The population size, N.
   (c) The sample size, n.
   (d) The sample mean.
   (e) The variance of the sample mean.
   (f) The largest value in the sample.
   (g) The population variance.
   (h) The estimated variance of the sample mean.

5. (Based on Rice 7.4) Two populations are surveyed with simple random sampling. A sample of size $n_1$ is used for population I, which has a population standard deviation of $\sigma_1$; a sample of size $n_2 = 3n_1$ is used for population II, which has a population standard deviation of $\sigma_2 = 2\sigma_1$.

(a) Ignoring the finite population correction, in which of the two samples would you expect the estimate of the population mean to be more accurate (i.e., smallest variance)? Provide a mathematical justification for your answer.

(b) For what ratio of $n_2/n_1$ would the estimates have equivalent accuracy (i.e., equivalent variances)?

(c) Verify this ratio via simulation, i.e., create populations I and II by simulating 1000 normal RVs for each with $\mu = 1$ and $\sigma_1 = 1$ and generate 1000 estimates of the population mean $\mu$ using random samples with $n_1 = 100$ and $n_2$ set to give the ratio you found in b). Plot the distributions of these estimates using a kernel density estimate (the distributions should look similar). Why won't these empirical distributions look identical?

6. (Based on Rice 7.10) True or false (and state why): If a sample from a population is large, a histogram of the values in the sample will be appropriately normal, even if the population is not normal? Verify your answer via simulation.

7. (Based on Rice 7.16) True or false? Justify your answers.

   (a) The center of a 95% confidence interval for the population mean is a random variable.

   (b) A 95% confidence interval for $\mu$ contains the sample mean with probability 0.95.

   (c) A 95% confidence interval contains 95% of the population.

   (d) Out of one hundred 95% confidence intervals for $\mu$, 95 will contain $\mu$.

8. (Optional; Based on Rice 7.19) This problem introduces the concept of a one-sided CI. Using the CLT, how should the constant k be chosen so that the interval $(\bar{X} - ks_{\bar{X}}, \infty)$ is a 90% CI for $\mu$, i.e., so that $P(\mu \geq \bar{X} + ks_{\bar{X}}) = 0.90$? This is called a one-sided CI.

   (a) Find k such that:

$$P(\mu \geq \bar{X} - ks_{\bar{X}}) = 0.9$$

   (b) How should k be chosen so that $(-\infty, \bar{X} + ks_{\bar{X}})$ is a 95% one-sided CI?