Assignment 4 - CSC 262 - Fall 2017 - Due November 16

1. Suppose the true mean and standard deviation of a population of measurements is $\mu = 76.5$ and $\sigma = 1.2$. A random sample of $n = 12$ is collected.

   (a) What is the standard deviation of the sample mean?

   (b) What is the probability that the sample mean is within 4 units of the true mean?

2. Suppose a random sample of $n = 6$ measurements is collected:

$$23.72, 18.95, 9.79, 12.04, 9.54.$$

   Assume that they are from a normally distributed population.

   (a) Construct a confidence level for the population mean using a 90% confidence level.

   (b) Do a hypothesis test of $H_o : \mu = 24$ against $H_a : \mu \neq 24$. Report a P-value. Do you reject H_o at significance level $\alpha = 0.05$? What about significance level $\alpha = 0.01$?

3. Suppose a population of measurements is claimed to be normally distributed with mean no larger that $\mu = 150$ and standard deviation $\sigma = 10$. To test this claim, a random sample of n components is to be collected to do a hypothesis test for $H_o : \mu = 150$ against $H_a : \mu > 150$.

   (a) Use `R` to draw a *power curve*, that is plot $Power(\mu) = 1 - \beta(\mu)$ as a function of μ over a suitable range of alternative hypotheses, say $\mu \in (150, 175)$ (use `seq(150,175,by=0.1)` to generate the values of μ for your plot). Do this for a Type I error of $\alpha = 0.05$. Superimpose on the same plot power curves for $n = 5, 10, 15, 20, 25, 30$. Label the appropriate axes μ and $Power(\mu)$. Include a horizontal line at level 0.05, labelled $\alpha = 0.05$. Also, indicate, using the `text()` function, the positions of the $n = 5$ and $n = 30$ curves.

   (b) Create a table giving the power for each combination of $n = 5, 10, \ldots, 30$ and

$$\mu = 155, 156, 157, 158, 159, 160.$$

   For each of these values of μ, give the minimum sample size (from those considered) required to attain a power of 80%.

4. We are given two independent samples from normally distributed populations. The data is summarized in the table below.

	Sample 1	Sample 2
$\bar{X}$	96.3	121.5
S	16.5	12.5
n	15	8

   (a) Construct a 98% confidence interval for the difference in means $\mu_2 - \mu_1$. Use two procedures, assuming i) equal and ii) unequal variances (that is, the pooled procedure and Welch's procedure). How do the confidence intervals differ?

   (b) Do a hypothesis test of $H_o : \mu_2 = \mu_1$ against $H_a : \mu_2 > \mu_1$. Again, use two procedures, assuming i) equal and ii) unequal variances. Report a P-value.

5. Two types of fish attractors, one made from vitrified clay pipes, and the other from cement blocks and brush, were used during 16 different time periods spanning four years at Lake Tohopekaliga, Florida. The following observations are of the average number of fish caught per fishing day.

Table 1: Fish caught per day, by time period and fish attractors.

Time Periods	Pipe	Brush
1	6.64	9.73
2	7.89	8.21
3	1.83	2.17
4	0.42	0.75
5	0.85	1.61
6	0.29	0.75
7	0.57	0.83
8	0.63	0.56
9	0.32	0.76
10	0.37	0.32
11	0.00	0.48
12	0.11	0.52
13	4.86	5.38
14	1.80	2.33
15	0.23	0.91
16	0.58	0.79

(a) Use the `R` `t.test()` function to do a paired t-test to determine whether or not one attractor is more effective. Use level $\alpha = 0.01$

(b) Repeat part (a), but assume the samples are independent. Use procedures for both equal and unequal variances. Does your conclusion change?

6. **[For Graduate Students]** Correlation between random variables can be modeled in the following way. Let X and ϵ be random variables with mean 0 and variance 1, and assume X and ϵ are independent. Then set, for constants $\beta_0, \beta_1, \beta_2$, a new random variable Y as follows:

$$Y = \beta_0 + \beta_1 X + \beta_2 \epsilon. \qquad (1)$$

(a) Derive an expression for the correlation between X and Y as a function of β_1 and β_2 (verify that this expression will not depend on β_0).

(b) Suppose we wish to simulate pairs of random variables (X, Y) such that $\mu_X = \mu_Y = 0$, $var(X) = var(Y) = 1$, and the correlation between X and Y is fixed at $\rho \in (-1, 1)$. We can do this by simulating X and ϵ, then using equation (1) to generate Y. To achieve this, what values must be used for $\beta_0, \beta_1, \beta_2$? Using `R`, generate four scatter plots from 1000 pairs of normally distributed random variables (X, Y) using this scheme, for $\rho = -0.5, 0, 0.5, 0.9$. Do an independent simulation for each of the four plots. Place all four scatter plots in one graphics window using the `par()` function. Indicate the relevant title for each plot, using the Greek font for ρ (consult `help(plotmath)` and the function `bquote()`).

In addition, for each simulation, summarize in a table the sample variances and correlation of X and Y. Compare these sample values to the theoretical values.