# STA261: Assignment 7

*Alex Stringer*

*February, 2018*

This assignment is not for credit. Complete the questions as preparation for quizzes and tests.

Suggested reading: Textbook sections 9.2, 9.3.

Do the following questions from textbook section 9.11 (page 362): 1a, 2, 5abdf, 9 (except the part about power),

1. Let $X_i \sim N(\mu_0, \sigma_0^2)$ be an IID random sample from a normal distribution with mean $\mu_0$ and variance $\sigma_0^2$. Derive a $1-\alpha$ confidence interval for

   (a) $\mu$, when $\sigma^2$ is *known*

   (b) $\mu$, when $\sigma^2$ is *uknown*

   (c) $\sigma^2$, when $\mu$ is *known*

   (d) $\sigma^2$, when $\mu$ is *unknown*

2. Let $X_i \sim Bern(\theta_0)$ be independent Bernoulli random variables.

   (a) Derive an approximate $1-\alpha$ confidence interval for $\theta$

   (b) Derive an approximate Hypothesis Test for testing $H_0 : \theta = \theta_0$

3. Suppose a pollster in an election with two candidates says that "based on a random sample of constituents asked who they prefer out of the two candidates, candidate A has a popularity of 47%, plus/minus 4 percentage points 19 times out of 20".

   (a) Use your answer from the previous question to write down a precise mathematical version of this statement.

   (b) Approximately how many constituents were sampled in this poll?

   (c) Could you increase the sample size so that an estimate of 47% resulted in an interval that did not contain 50%? If so, what would the required sample size be?

4. Prove Theorem B of section 6.3 in the textbook; that if $X_i \sim N(\mu_0, \sigma_0^2)$ are IID then

$$\frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{n-1}$$

   where

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Like the textbook, you may begin with (you don't have to prove) the fact that

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu_0}{\sigma}\right)^2 \sim \chi_n^2$$

5. *t-distribution.* Let $Z \sim N(0,1)$, $U \sim \chi_k^2$, and $Z \perp U$. In this question you will derive the density of

$$T = \frac{Z}{\sqrt{U/k}}$$

   (a) Use the change of variables formula (textbook, section 2.3, page 62, proposition B) to show that the density of $V = \sqrt{U/k}$ is

   $$f_V(v) = \frac{k^{k/2}}{\Gamma(k/2)2^{k/2-1}} v^{k-1} e^{-kv^2/2}$$

   Note that $U > 0$ with probability 1, so the transformation here is invertible.

   (b) Recall the formula for the density of the *quotient* of two independent random variables (textbook section 3.6.1, page 98): if $Z = Y/X$, and $X > 0$ with probability 1, then

   $$f_Z(z) = \int_0^\infty x f_X(x) f_Y(xz) dx$$

   (I added the condition that $X > 0$ with probability one, so $f_X(x) = 0$ if $x < 0$, to simplify the formula for use in this problem). Use this and the independence of $Z$ and $U$ to show that the density of $T = Z/V$ is given by

   $$f_T(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma(k/2)}\left(1+\frac{t^2}{k}\right)^{-\frac{k+1}{2}}$$

   You can use the fact that if $X \sim Gamma(\alpha, \beta)$,

   $$f_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$$

   which lets you evaluate expressions of the form

   $$\int_0^\infty x^{\alpha-1} e^{-x/\beta} dx = \Gamma(\alpha)\beta^\alpha$$

6. Show the equivalence between confidence intervals and hypothesis tests for normally distributed samples. That is, for a hypothesis test with rejection region of the form

$$R_\alpha(T) = (-\infty, -z_{1-\alpha/2}) \cup (z_{1-\alpha/2}, \infty)$$

and test statistic

$$T = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$$

show that $T \in R_\alpha(T)$ if and only if $\mu_0$ does not lie in the corresponding $1-\alpha$ confidence interval.

7. Let $X_i \sim N(\mu_0, \sigma_0^2)$ be an IID sample from a normal distribution with both parameters unknown.

(a) Use the distributional result derived in class and earlier on this assignment,

$$\frac{(n-1)s^2}{\sigma_0^2} \sim \chi^2_{n-1}$$

to derive a $1-\alpha$ confidence interval for $\sigma^2$. Note that the $\chi^2$ distribution is *not* symmetric, so your answer will include both quantiles $\chi^2_{n-1,\alpha/2}$ and $\chi^2_{n-1,1-\alpha/2}$. Note also that you already did this in question 1, so just write it down again.

(b) For the following sample sizes and sample variances, compute a 95% confidence interval for $\sigma^2$. My answers are included. You can use the `qchisq` function in `R` to compute the necessary quantiles, or look them up in a table/online calculator.

(i) $n = 10, s^2 = 2$. My answer: 0.95, 6.67

(ii) $n = 100, s^2 = 2$. My answer: 1.54, 2.7

(iii) $n = 20, s^2 = 100$. My answer: 57.83, 213.33

8. For the following sample sizes, find a 95% confidence interval for $\mu$ using both the $t$-distribution and the normal distribution, for $\bar{X} = 2.7$ and $s^2 = 4$. When doing the normal distribution ones, just use $s^2$ for $\sigma^2$- the purpose of this question is to show you how properly accounting for the uncertainty in estimating $\sigma^2$ changes the width of the confidence interval, and how this is affected by sample size. You can use the `qt` and `qnorm` functions in `R` to compute the quantiles.

   (a) $n = 3$. My answers: Normal (0.44, 4.96), t (-2.27, 7.67)

   (b) $n = 10$. My answers: Normal (1.46, 3.94), t (1.27, 4.13)

   (c) $n = 30$. My answers: Normal (1.98, 3.42), t (1.95, 3.45)

   (d) $n = 100$. My answers: Normal (2.31, 3.09), t (2.3, 3.1)

9. For each of the above scenarios, test the hypothesis that $\mu_0 = 2$ using both the Normal and $t$-distributions. Report your p-value, whether you reject or not, and whether your result agrees with the corresponding confidence interval. My p-values are

   (a) Z: 0.54; t: 0.61

   (b) Z: 0.27; t: 0.3

   (c) Z: 0.06; t: 0.07

   (d) Z: 0; t: 0

10. The professor who taught me introductory statistics (see youtube channel: jbstatistics) went out and bought $n = 16$ boxes of Raisin Bran and measured their weights, with the goal of testing Kellogs' claim that the mean weight of cereal present in boxes of Raisin Bran is 755g (family size). I don't have these data but here are some synthetic datapoints representing the weights (in grams) of some randomly purchased boxes of family size raisin bran:

$$782.1, 765, 708.9, 768.6, 751.5, 732, 763.3, 766.5, 768.7, 749.2, 750.5, 748.4, 781.8, 779, 775.2, 746.9$$

Answer the following questions; my answers are included.

(a) Find a 95% confidence interval for the mean weight of boxes of raisin bran sampled in this way (My answer: 747.77, 769.43)

(b) Test the hypothesis that the mean weight of boxes of raisin bran sampled in this way is actually 755g, at the 0.05 significance level (My answer: $p = 0.46$)

(c) How many boxes of raisin bran would you need to buy such that observing a sample with this same mean and standard deviation would cause you to reject the null hypothesis? You won't get a closed form answer for $n$. Show that this sample size must satisfy

$$\frac{t_{n-1,1-\alpha/2}}{\sqrt{n}} < 0.1841496$$

You can use `R` or Excel to find $n$ numerically by brute force; I got $n = 116$