170S Week 9 Discussion Notes

Colin Ni

March 2, 2025

This week we will take it easy and do some problems from the two homework assignments due on Sunday, namely some even-numbered ones because the odd-numbered ones have answers in the back of the book. As usual, I highly recommend that when you write your solution down, you do the problem entirely on your own (except with the help of a calculator and the tables).

Exercise 7.4-2. Let X equal the excess weight of soap in a "1000-gram" bottle. Assume that the distribution of X is $\mathrm{Norm}(\mu, 169)$. What sample size is required so that we have 95% confidence that the maximum error of the estimate of μ is 1.5?

Solution. Given a sample $X_1, \ldots, X_n$, we have

$$\frac{\bar{X} - \mu}{\sqrt{169/n}} \sim \mathrm{Norm}(0, 1).$$

A 95% confidence interval for this quantity is $(-z_{0.025}, z_{0.025})$, so a 95% confidence interval for the mean μ is

$$(\bar{X} - z_{0.025}\sqrt{169/n}, \bar{X} + z_{0.025}\sqrt{169/n}).$$

The *maximum error of the estimate* is by definition (see top of page 331) half the width of this interval, *i.e.* $z_{0.025}\sqrt{169/n}$, and in order for

$$z_{0.025}\sqrt{169/n} \leq 1.5,$$

we must have

$$288.55 \approx \left(\frac{z_{0.025}\sqrt{169}}{1.5}\right)^2 \leq n.$$

Thus we need a sample size of at least 289. □

Exercise 7.5-4. Let m denote the median weight of "80-pound" bags of water softener pellets. Use the following random sample of $n = 14$ weights to find an approximate 95% confidence interval for m:

$$80.51, 80.28, 80.16, 80.59, 80.40, 80.35, 80.56,$$
$$80.32, 80.38, 80.28, 80.27, 80.53, 80.27, 80.32.$$

(a) Find a 94.26% confidence interval for m.

(b) The interval $(x_{(6)}, x_{(12)})$ could serve as a confidence interval for $\pi_{0.6}$. What is its confidence coefficient?

Solution. For (a), recall that $(x_{(i)}, x_{(n+1-i)})$ is a distribution-free confidence interval for the median m with confidence coefficient

$$\begin{aligned}\mathbb{P}(m \in (x_{(i)}, x_{(n+1-i)})) &= \sum_{k=i}^{n-i} \mathbb{P}\left(\mathrm{Binom}\left(n, \frac{1}{2}\right) = k\right) \\ &= \sum_{k=i}^{n-i} \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k}\end{aligned}$$

by assuming that our sample median is the 50th percentile (so that samples fall to the left or right of the sample median with a 50/50 chance). Table 7.5-1 gives for some values of n a value of i such that the confidence coefficient of this interval is close to 0.95. For $n = 14$, it gives $i = 4$ so that the confidence interval is

$$(x_{(4)}, x_{(11)}) = (80.28, 80.51),$$

and it gives that the confidence coefficient is 0.9426, which matches the one in the exercise.

For (b), now we are interested in when samples fall to the left or right of the sample percentile $\pi_{0.6}$, and these happen with probability 0.6 and 0.4. Thus

$$\mathbb{P}(\pi_{0.6} \in (x_{(6)}, x_{(12)}))) = \sum_{k=6}^{11} \binom{n}{k} \cdot 0.6^k \cdot 0.4^{n-k} \approx 0.9019. \qquad \square$$

Exercise 8.2-4. Among the data collected for the World Health Organization air quality monitoring project is a measure of suspended particles, in $\mu\mathrm{g/m}^3$. Let X and Y equal the concentration of suspended particles in $\mu\mathrm{g/m}^3$ in the city centers (commercial districts) of Melbourne and Houston, respectively. Using $n_X = 13$ observations of X and $n_Y = 16$ observations of Y, we shall test $H_0\colon \mu_X = \mu_Y$ against $H_1\colon \mu_X < \mu_Y$.

(a) Define the test statistic and critical region, assuming that the variances are equal. Let $\alpha = 0.05$.

(b) If $\bar{x} = 72.9$, $s_X = 25.6$, $\bar{y} = 81.7$, and $s_Y = 28.3$, calculate the value of the test statistic and state your conclusion.

(c) Give bounds for the p-value of this test.

Solution. We will assume the concentrations are normal. Recall that since the variances are equal, we have that

$$\frac{(\bar{x} - \bar{y}) - (\mu_X - \mu_Y)}{s_P\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \sim t(n_X + n_Y - 2)$$

where

$$s_P = \sqrt{\frac{(n_X - 1)s_X^2 + (n_Y - 1)s_Y^2}{n_X + n_Y - 2}}$$

follows a t-distribution with $n_X + n_Y - 2$ degrees of freedom. Thus

$$\left(\bar{x} - \bar{y} - t_\alpha(n_X + n_Y - 2)s_P\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}, \qquad \infty \right)$$

is a right-tail confidence interval for $\mu_X - \mu_Y$ with confidence coefficient $1 - \alpha$.

For (a), let us translate this into a hypothesis test. Let us choose our test statistic to be $\bar{x} - \bar{y}$. The critical region, *i.e.* the region where we reject the null hypothesis, is where

$$\bar{x} - \bar{y} \in \left(-\infty, \qquad -t_\alpha(n_X + n_Y - 2)s_P\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}} \right).$$

For (b), we have $t_{0.025}(n_X + n_Y - 2) \approx 2.052$, so the critical region is $(-\infty, -20.7900)$. Since our test statistic is $\bar{x} - \bar{y} = -8.8$ and is not in the critical region, we do not reject the null hypothesis. $\square$

Remark. When you do this problem yourself, in particular part (b), I recommend using the test statistic

$$\frac{\bar{x} - \bar{y}}{t_\alpha(n_X + n_Y - 2)s_P\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}}$$

instead which follows a t-distribution. See table 8.2-1 to check your work on determining a confidence interval and critical region. It is important that you get used to using various test statistics not only because it builds intuition but also because on an exam you may be asked to use a specific one.