Steve Sawin
Statistics 3352

Test II (Solutions)

You may use two pages of notes, calculator and computer. Please sign your name below indicating that you have not offered or received help on this test to or from any person, nor used any resource other than your notes, me, and the Excel files on my website.

---

1. [8pt] Use the method of moments to find an estimator for σ^2 for a variable X.

$$\sigma^2 = \mu_2' - \mu^2$$

so the best estimator is

$$\frac{\sum_i X_i^2}{n} - \overline{X}^2.$$

2. [10pt] I sample 8 independent individuals from a population with a normal variable X, and compute $\overline{X}$ and S. What is the probability that

$$\overline{X} - \mu > S/4?$$

Helpful Hint: For T a t-dist with $n-1$ degrees of freedom, in Excel and TI-84 respectively $P(T > t) = \text{TDIST}(t, n-1, 1) = \text{tcdf}(t, 99999, n-1)$.

$$\begin{aligned} P(\overline{X} - \mu > S/4) &= P\left(\frac{\overline{X} - \mu}{S/\sqrt{8}}\right) > \frac{\sqrt{8}}{4} \\ &= P\left(T_7 > \frac{\sqrt{8}}{4}\right) = \text{TDIST}(\frac{\sqrt{8}}{4}, 7, 1) = 25.1\% \end{aligned}$$

where T_7 is a variable with a t distribution with 7 degrees of freedom.

3. Let $f(x)$ be the pdf for the normal distribution with $\sigma = 1$, (I am setting $\sigma = 1$ to make the calculation as simple as possible, don't let it confuse you).

   (a) [4pt] What is $f(x)$?

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2}}$$

   (b) [10pt] Find

$$\frac{1}{nE\left[\left(\frac{\partial \ln[f(x)]}{\partial \mu}\right)^2\right]}?$$

$$\ln[f(x)] = -\frac{(x-\mu)^2}{2} - \ln\left(\sqrt{2\pi}\right).$$

$$\frac{\partial \ln[f(x)]}{\partial \mu} = x - \mu.$$

$$\left(\frac{\partial \ln[f(x)]}{\partial \mu}\right)^2 = (x-\mu)^2.$$

$$E\left[\left(\frac{\partial \ln[f(x)]}{\partial \mu}\right)^2\right] = E\left[(x-\mu)^2\right] = \sigma^2 = 1.$$

$$\frac{1}{nE\left[\left(\frac{\partial \ln[f(x)]}{\partial \mu}\right)^2\right]} = \frac{1}{n}$$

(c) [7pt] Show $\overline{X}$ is a *minimum variance, unbiased* estimator for μ if X is a normal variable with $\sigma = 1$. By Cramer-Rao and the previous part it is enough to show that (unbiased)

$$E\left(\overline{X}\right) = \mu$$

which we know for any sample of any variable and (minimum variance)

$$\text{Var}\left(\overline{X}\right) = \frac{1}{n}$$

which we know since the sample is independent

$$\text{Var}\left(\overline{X}\right) = \frac{\text{Var}(X)}{n} = \frac{\sigma^2}{n} = \frac{1}{n}.$$

4. You want to know how residents of the town of Fairfield feel about FU students, so you stop 60 people coming out of the Seagrape on Saturday night and ask them, and find that 48 of them like Fairfield students.

(a) [10pt] Find a 90% confidence interval for the proportion who like Fairfield students and express it in a sentence that explicitly includes the parameter, population and variable. (answer is a sentence with some numbers in it)

$$\hat{p} = 48/60 = .8 \qquad \sigma_{\hat{p}} = \sqrt{.8 * .2/60} = .052.$$

$$.8 \pm 1.645 * .052 = .8 \pm 0.085$$

We are 90% confident that the proportion of all Fairfield residence who like FU students is between 71.5% and 88.5%.

(b) [5pt] What does the 90% above mean? Specifically, interpret the confidence interval by identifying a set of things (probability space) and a statement about them (event) that is is true for 90% of them. (answer is a short sentence) 90% of all samples of size 50 will give a 90% confidence interval which contains the true proportion of all Fairfield residents who like FU students.

(c) [6pt] Check each assumption of this procedure. (answer is a few words explaining why each is or is not met)
SRS: Clearly **not** met: people coming out of the Grape at night are more likely to be FU students, more likely to be young, and more likely to like FU students.

Large Pop: Met. Surely there are more than 800 residents in the town of Fairfield.

Normality: Rule of 15. **Not met.** there are 12 no answers, which is less than 15.

(d) [5pt] Identify a (pretty dramatic) source of sampling bias in the above study. People coming out of the Grape on Saturday night are going to be overwhelmingly FU students and therefore (hopefully) more inclined to like FU students (even if some townies go to thr Grape, they are presumably going to be those that like Stags).

[12pt] A SRS of 50 Millenials asks how often in a week each used the "laughing-cry" emoji, and found $\overline{X} = 13.2$ and $S = 9$. A study asking 42 Gen Zers the same question found $\overline{X} = 8.9$ and $S = 7.1$. Find the 92% confidence interval for how much more Millenials use it than Gen-Z on average, stated in a complete English sentence. The 92% confidence interval for the amount by which Millenials exceed Gen-Z in their average use of the emoji is 4.30 ± 2.97 uses per week.

5. A simple random sample of mathematics textbooks had the following number of pages:

580 1252 809 413 651 588 169 537 763 816.

(a) [5pt] Find the sample mean $\overline{X}$ and the sample standard deviation S. To check you typed the info in correctly, I got $\overline{(}X) = 65?.80$, Where ? is a digit you still need to figure out.

$$\overline{(}X) = 657.80$$
$$S = 286.05$$

(b) [12pt] Give the 96% confidence interval for the average number of pages of all math books in a complete English sentence. The 96% confidence interval for mean number of pages of all math books is 657.80 ± 216.96.

(c) [6pt] Check all three assumptions. It says it is an SRS. There are surely more than $20 * 10 = 200$ math bo0oks in the world so the first two are met. Since $n = 10$ we would need to know that mathbook pages follow a normal distribution, and I do not know that!

out of 100 points