Math Stat, solutions to HW4

Practice problems.

Book 18. The general form with known variance is

$$\mu = \bar{X} \pm z\frac{\sigma}{\sqrt{n}} \quad (q)$$

where $\Phi(z) = (1+q)/2$. We have $\sigma = 1, n = 5, \bar{X} = 1000.3$ and with $q = 0.95$, we get $(1+q)/2 = 0.975$ and Table A2 gives $z = 1.96$. Hence

$$\mu = 1000.3 \pm 1.96\frac{1}{\sqrt{5}} = 1000.3 \pm 0.88 \quad (0.95)$$

Book 1. Since X is a sum of r squares of independent standard normal random variables and Y is a sum of another s squares of independent standard normal random variables, $X+Y$ is a sum of $r+s$ squares of independent standard normal random variables and hence $X+Y \sim \chi^2_{r+s}$.

Book 9. The general form with unknown variance is

$$\mu = \bar{X} \pm t\frac{s}{\sqrt{n}} \quad (q)$$

where $F_{n-1}(t) = (1+q)/2$. We have $\bar{X} = 0.11, n = 7$ and get

$$s^2 = \frac{1}{n-1}\left(\sum_{k=1}^{n} X_k^2 - n\bar{X}^2\right) = \frac{1}{6}(0.06^2 + ... + 0.21^2 - 7\cdot 0.11^2) = 0.0027$$

so that $s = \sqrt{0.0027} = 0.05$. With $q = 0.95$, $(1+q)/2 = 0.975$ and Table A3 gives $t = 2.45$ (DF$= n-1 = 6$). We get

$$\mu = 0.11 \pm 2.45\frac{0.05}{\sqrt{7}} = 0.11 \pm 0.05 \quad (0.95)$$

1(a) Consider the sample of the differences "after minus before." This sample is $17, 2, -1, -14, -15$ which has $\bar{X} = -2.2$ and $s = 13.1$. With $q = 0.95$, we get $t = 2.78$ (Table A3 with DF$= n-1 = 4$ and $(1+q)/2 = 0.975$). The confidence interval becomes

$$\mu = -2.2 \pm 2.78\frac{13.1}{\sqrt{5}} = -2.2 \pm 16.3 \ (0.95)$$

The "before minus after" value is of the form $X - Y$ where both X and Y are normal and as linear combinations of normals are normal, we conclude that $X - Y$ has a normal distribution.

(b) No. The confidence interval contains 0 which is the point where there is no difference. We would need the interval to be entirely above or entirely below 0 for there to be a significant effect.

2. The general form of the t distribution is

$$\frac{Z}{\sqrt{\frac{Y}{r}}}$$

where the numerator and denominator are independent, $Z \sim N(0,1)$, and $Y \sim \chi^2_r$, the distribution of the sum of squares of r independent $N(0,1)$. Since we can write T as

$$T = \frac{Z_1}{\sqrt{\frac{Z_2^2 + Z_3^2 + Z_4^2 + Z_5^2}{4}}}$$

where the numerator is $N(0,1)$, the denominator is of the form "$\chi^2_4/4$," and the numerator and denominator are independent (contain different Z_k), we conclude that $T \sim t_4$.

Turn-in problems

1(a) Let $\widehat{\theta} = \frac{n+1}{n}X_{(n)}$ and note that

$$\frac{\widehat{\theta}}{\theta} \leq a \Leftrightarrow \theta \geq \frac{\widehat{\theta}}{a}$$

where the distribution of $\widehat{\theta}/\theta$ is known:

$$P\left(\frac{\widehat{\theta}}{\theta} \leq a\right) = \left(\frac{n}{n+1}\right)^n a^n$$

as shown in class. As this is the case when our interval misses θ, we want the probability to be $1-q$ which gives

$$a = \frac{n+1}{n}(1-q)^{1/n}$$

and the confidence interval

$$\theta \le \widehat{\theta}\ \frac{n+1}{n}(1-q)^{1/n} \quad (q)$$

(b) If we remove the lower bound in a two-sided interval the th confidence level q, the resulting one-sided interval no longer has confidence level a but $(1+q)/2$.

2(a) Not a t distribution because the numerator and denominator are not independent (Z_1 appears in both).

$$\frac{Z_2}{\sqrt{\dfrac{Z_3^2+Z_4^2+Z_5^2+Z_6^2}{4}}}$$

(b) This is a t distribution with 1 degree of freedom since

$$|Z_2| = \sqrt{\frac{Z_2^2}{1}}$$

(c) This is a t distribution with 2 degrees of freedom because we can rewrite it as

$$\frac{\dfrac{Z_1+Z_2}{\sqrt{2}}}{\sqrt{\dfrac{Z_3^2+Z_4^2}{2}}}$$

where the numerator is $N(0,1)$ because the sum of normals is normal and $E[Z_1+Z_2] = E[Z_1]+E[Z_2] = 0$ and

$$\mathrm{Var}\left[\frac{Z_1+Z_2}{\sqrt{2}}\right] = \frac{1}{2}(\mathrm{Var}[Z_1]+\mathrm{Var}[Z_2]) = 1$$

3(a) This is a normal sample with unknown variance. We have $n = 7$, $\bar{X} = 0.11$ and $s = 0.05$. With $q = 0.95$, we get $t = 2.45$ (Table A2, p.853) with DF$= n - 1 = 6$ and $1 - (1 + q)/2 = 0.025$. The confidence interval becomes

$$\mu = 0.11 \pm 2.45 \frac{0.05}{\sqrt{7}} = 0.11 \pm 0.05 \ (0.95)$$

(b) No, because the interval contains 0.10 we cannot be "95% certain" that μ is above 0.10. If our interval was entirely above 0.10, we could claim such 95% certainty.

In reality, it makes more sense to do a one-sided lower-bonded confidence interval because we are not interesting in how high μ is, only whether it is above 0.10. This will give a bound that is higher than the lower bound in the two-sided interval so there may be cases when the one-sided and two-sided intervals lead to different conclusions.