1. Let N and N' be independent Normal$(0,1)$ random variables. Find the probabilities that (a) $2N > N'+1$ and (b) $(N-N')^2+(N+N')^2 > 4$.

   Solution:

   (a) $2N$ is Normal$(0,2)$. As N' is independent of $2N$, $2N-N'$ is Normal$(0,\sqrt{5})$, so

   $$\begin{aligned} \mathrm{P}(2N > N'+1) &= \mathrm{P}(2N - N' > 1) \\ &= \mathrm{P}(\text{Normal}(0,\sqrt{5}) > 1) \\ &= \mathrm{P}(\text{Normal}(0,1) > 1/\sqrt{5}) \\ &\approx 0.32736. \end{aligned}$$

   (b) $N^2 + N'^2$ is a $\chi^2(2)$ random variable, so

   $$\begin{aligned} \mathrm{P}((N-N')^2+(N+N')^2 > 4) &= \mathrm{P}(2N^2+2N'^2 > 4) \\ &= \mathrm{P}(N^2+N'^2 > 2) \\ &= \mathrm{P}(\chi^2(2) > 2) \\ &\approx 0.36788. \end{aligned}$$

2. The midterm test scores of six random students are 81, 84 ,83, 73, 76, 83.

   (a) What is the sample variance?

   Solution: The sample mean is $\overline{X} = \frac{81+84+83+73+76+83}{6} = 80$, so the sample variance is

   $$V = \frac{(81-80)^2+(84-80)^2+(83-80)^2+(73-80)^2+(76-80)^2+(83-80)^2}{6} = \frac{50}{3}.$$

   (b) Assuming their scores are independent Normal(μ,σ) random variables. Give as large a value for $\hat{\Theta}_-$ as you can so that $(\hat{\Theta}_-, 10)$ is a 95% confidence interval for σ.

   Solution: The adjusted sample variance is $S^2 = 50 \cdot 6/(6-1) = 20$. The confidence interval formula for the standard deviation of a normal random variable is $(\sqrt{(n-1)S^2/z_+}, \sqrt{(n-1)S^2/z_-})$, where z_- and z_+ should be chosen so that

   $$\mathrm{P}(z_- \le \chi^2(n-1) \le z_+) = 0.95.$$

   For the right bound to equal 10 we should choose $z_- = 1$. As $\mathrm{P}(\chi^2 < z_-) \approx 0.03743$ we are looking for z_+ with $\mathrm{P}(\chi^2(n-1) \le z_+) \approx 0.95 + 0.03743 \approx 0.98743$, which gives $z_+ \approx 14.53$ and $\hat{\Theta}_- = \sqrt{(n-1)S^2/z_+} \approx 2.6234$.

3. A food processing company packages honey in glass jars. The volume of honey in millilitre in a random jar is a Normal(μ, σ) random variable. 5 random jars are picked and the volume of honey inside them in millilitre are 108, 101, 103, 109 and 104.

   (a) Suppose μ is unknown and σ is known to be 5. Give a 95% confidence interval for μ.

   Solution: The sample mean is $\overline{X} = \frac{108+101+103+109+104}{5} = 105$, so the confidence interval is $(\overline{X} - z\sigma/\sqrt{n}, \overline{X} - z\sigma/\sqrt{n}) \approx (100.64, 109.36)$, where $z \approx 1.96$ is chosen so that $\mathrm{P}(-z \le \text{Normal}(0,1) \le z) = 95\%$.

(b) Suppose μ and σ are both unknown. Give a 95% confidence interval for μ.

Solution: The adjusted sample variance is:

$$S^2 = \frac{(108-105)^2 + (101-105)^2 + (103-105)^2 + (109-105)^2 + (104-105)^2}{5-1} = \frac{23}{2}$$

The confidence interval is now of the form $(\overline{X} - zS/\sqrt{n}, \overline{X} + zS/\sqrt{n})$, where z is chosen so that $\mathrm{P}(-z \leq t(4) \leq z) = 94\%$, which gives $z \approx 2.78$ and the confidence interval $(100.29, 109.71)$.

(c) Suppose μ and σ are both unknown. Give a 95% prediction interval for the next sample.

Solution: The prediction interval is of the form $(\overline{X} - zS\sqrt{1+1/n}, \overline{X} + zS\sqrt{1+1/n})$ with the same z as in part (b), giving the answer $(94.67, 115.33)$.

4. In this question you will study prediction intervals for normal random variables with known and unknown standard deviation σ.

(a) If $X_1, \dots, X_n, X_{n+1}$ are independent Normal(μ, σ) random variables, what kind of random variable is $X_{n+1} - (X_1 + \cdots + X_n)/n$?

Solution: By the algebra of independent normals, $X_1 + X_2 + \cdots + X_n$ is Normal$(n\mu, \sqrt{n}\sigma)$, $\overline{X} = (X_1 + X_2 + \cdots + X_n)/n$ is Normal$(\mu, \sigma/\sqrt{n})$, and $X_{n+1} - \overline{X}$ is Normal$(\mu - \mu, \sqrt{\sigma^2 + \sigma^2/n})$ = Normal$(0, \sqrt{1+1/n} \cdot \sigma)$.

(b) Based on part (a) come up with a prediction interval for the next Normal(μ, σ) sample for unknown μ and known σ.

Solution: From part (a), $(X_{n+1} - \overline{X})/(\sqrt{1+1/n} \cdot \sigma)$ is a Normal$(0,1)$ random variable, so

$$\mathrm{P}(z_- \leq (X_{n+1} - \overline{X})/(\sqrt{1+1/n} \cdot \sigma) \leq z_+) = \mathrm{P}(z_- \leq \text{Normal}(0,1) \leq z_+).$$

The left-hand event is $\overline{X} + z_-\sqrt{1+1/n} \cdot \sigma) \leq X_{n+1} \leq \overline{X} + z_+\sqrt{1+1/n} \cdot \sigma)$, which gives the prediction interval

$$\big(\overline{X} + z_-\sqrt{1+1/n} \cdot \sigma, \overline{X} + z_+\sqrt{1+1/n} \cdot \sigma)\big),$$

where z_- and z_+ are chosen so that the confidence level is $\mathrm{P}(z_- \leq \text{Normal}(0,1) \leq z_+)$.

(c) Assuming $\sigma = 5$, give a 95% prediction interval for the next sample from the following data: $37.1, 44.0, 49.3, 49.7, 37.9, 65.5, 63.7, 44.1, 58.3, 60.9$.

Solution: There are $n = 10$ samples with sample mean

$$\overline{X} = \frac{37.1 + 44.0 + 49.3 + 49.7 + 37.9 + 65.5 + 63.7 + 44.1 + 58.3 + 60.9}{10} = 51.05$$

Choosing $z_+ = -z_- = 1.96$ gives the 95%-confidence interval $(40.77, 61.33)$.

(d) Repeat part (c) for unknown σ. How do the confidence intervals in parts (c) and (d) compare?

Solution: The adjusted sample variance is $S^2 = 109.87$, so $S \approx 10.48$. Now the confidence interval is $(\overline{X} + z_-\sqrt{1-1/n}S, \overline{X} + z_+\sqrt{1-1/n}S)$, where z_- and z_+ should be chosen so that $\mathrm{P}(z_- \leq t(9) \leq z_+) = 95\%$. If we want them equal in magnitude we get $z_+ = -z_- \approx 2.262$. The confidence interval is now $(26.21, 75.89)$.

This confidence interval is (much) wider than the one in part (c) for two reasons: First, the sample standard deviation S is larger than the actual one σ. Second, the z-values for the $t(9)$ random variable are larger than the ones for the standard normal. Here the effect of the standard deviation is much larger ($S/\sigma \approx 2.24$ while $z_c/z_d \approx 1.15$). The discrepancy between S and σ is quite large so it might be worth investigating if the value for σ provided in part (c) could be wrong (assuming the data is correct). In fact, the lower 95% confidence bound for σ is about 7.64, so based on this data we can claim that σ is not equal to 5 with 95% confidence.