EXERCISE 2

We consider the question of whether the mean height of adult males in Great Britain has increased in the past 130 years or so? The current average height in Great Britain for adult males can be taken as 5’ 10” or 70 inches. In 1885 Sir Francis Galton developed a data set on heights for 465 adult males. We consider the complete sample and a randomly selected subset of 25 observations.

1. Let \( x_i \) denote the value of height for the \( i \)-th individual in the sample. Suppose the 1885 population mean is \( \mu \) and the variance is \( \sigma^2 \). Further suppose that \( x_i \sim i.i.d. N(\mu, \sigma^2) \).
   
   (a) Obtain the population mean of the sample average \( \bar{x}_n = \frac{1}{n} \sum_{i=1}^{n} x_i \).
   (b) Obtain the population variance of the sample mean \( \bar{x}_n \).
   (c) What is the distribution of the sample mean?
   (d) What is the distribution of \( (\bar{x}_n - \mu)/\sqrt{\sigma^2/n} \)?
   (e) How might we estimate the variance \( \sigma^2 \)?

2. Suppose the variance is known. We seek to test whether the mean height in 1885 is the same as today or \( H_1 : \mu = 70 = \mu_0 \) against the alternative \( H_1 : \mu = \mu_1 \neq 70 = \mu_0 \).
   
   (a) What will be the distribution of \( (\bar{x}_n - \mu_0)/\sqrt{\sigma^2/n} \), under the null hypothesis?
   (b) What will be the distribution of \( (\bar{x}_n - \mu)/\sqrt{\sigma^2/n} \), under the alternative hypothesis?
   (c) How will this distribution change for \( (\bar{x}_n - \mu)/\sqrt{s^2/n} \), under the null hypothesis, where \( s^2 \) is the estimated variance obtained in 1(e) above?
   (d) Discuss how we might use the result in (c) to conduct a test of the null hypothesis.
   (e) What does this test tell about choosing between the null and alternative hypotheses?

3. Our subsample of heights is: (65 70 69 64.7 73 71.2 69.2 66 64 73 68 68 72 69.5 69 69 67.7 66 73 76.5 66.5 68.5 66.5 69 67).
   
   (a) Conduct a two-sided test of the null hypothesis using this data. Justify your choice of critical value.
   (b) What does it mean to reject or not with this test?
   (c) Conduct a one-sided test of the null. Why is the outcome different for this test?
   (d) How can we know whether to conduct a one- or two-sided test?
4. Now consider the complete sample of 465 observations. Suppose $\sum_{i=1}^{n} x_i = 32,191$ and $\sum_{i=1}^{n} x_i^2 = 2,231,786$.

(a) Calculate the sample mean and variance for the complete sample.

(b) Compare these results with those for the subsample. Why are they the same or different?

(c) Conduct a two-sided test of the null hypothesis using the complete sample. Where do you find appropriate critical values for this test?

(d) Compare your results to those from the subsample. Do we expect the results to change in this fashion?