9 {R continued, unbiased-ness assumptions, small sample hypothesis test }
a. Given you NHIS sample you initially downloaded, drop observations where height is missing, restrict the sample to men and draw a sample of size 25. (Please show your sample and your code)

```
men_samp <- sample(NHIS.male$height,25)
men_samp
[1] 65 72 66 75 70 68 72 69 72 70 69 63 73 72 71 69 73 71 72 74 76 69 69 71 67
```

b. Compute the mean and the standard error of the mean for this sample.
(Please show your results and your code).

```
mean(men_samp)
[1] 70.32   #(your numbers will be different)

var(men_samp)^.5/25^.5
[1] 0.6102459    #(your numbers will be different)
```

c. Do you believe your estimate from part b is an unbiased estimate of the average weight of an American male? Why or why not?

Yes because I have drawn a random sample from the population (A1) and I believe that people are honestly reporting their heights (A2) so the two assumptions under which the sample mean is an unbiased estimate of the population mean are met.

d. Use your sample from parts a & b to conduct a significance test at the 5 percent level to test the null hypothesis that the average height of men in the population is 68.5 inches.

d1. State your null and alternative hypothesis,

$NULL, \quad H_0: \mu_{height} = 68.5$

$Alt \quad H_1: \mu_{height} \neq 68.5$

d2. Estimate the t-stat:

$$t_{\hat{\mu}_h} = \frac{\hat{\mu}_h - \mu_0}{se(\hat{\mu}_h)} = \frac{70.32 - 68.5}{0.6102459} = 2.9824$$

d3. Conduct your significance test (under the assumption that weight is distributed normally) - Which distribution did you check your t-statistic against and why?

$CV(dof = 24, 2tail, 5\%\ sig) = 2.064 \quad \sim use\ student's\ t - table$

In this case the degrees of freedom is 24 (n-1). When we look up the cut off for a two sided 5 percent significance test on the table for the Student's t-distribution we get a cut o of 2.064. I checked it against the Student's t distribution because when the variable we are estimating the mean for is normally distributed the t-statistic follows the Students t-distribution (the sample size is less than 30).

d4. What do you conclude about the null hypothesis?
Since $2.9824 > 2.064$

$\Rightarrow |t - stat| > C.V.$

We conclude that there is compelling evidence that the null is false. I conclude (at the 5% significance level, at the 95% confidence level) that the average male probably isn't 68.5 inches high.
(Note, although my example rejects the NULL, I set this up such that most students will likely fail to reject the NULL, implying that "In this case I conclude that there is no compelling evidence to reject the null hypothesis that the mean height in the population is 68.5 pounds.")

e. Calculate the p-value (two-tailed, approximate value with an inequality is fine).

[giveaway question, we will likely not have covered p-values by this point in lecture. but I trust student's ability to either remember their intro to stats class, or be capable of using google to learn how to find a p-value]

With a t-stat of 2.9824, and a two-tailed test, and 24 degrees of freedom, turning to a t-table, we find that the p-value is less than 0.01.