1. Suppose that the US intelligence quotients (IQs) are normally distributed with mean 100 and standard deviation 16. What's IQ score represents the 5^{th} percentile? (Explain your calculation.)

The fifth percentile for the standard normal is -1.645. Therefore, the fifth percentile for this population is

100 - 1.645 * 16 = 73.68

2. Consider the previous question. Note that 116 is the 84^{th} percentile from this distribution. Suppose now that $1,000$ subjects are drawn at random from this population. Use the central limit theorem to write the probability that less than 82% of the sample has an IQ below 116 as a standard normal probability. Note, you do not need to solve for the final number. (Show your work.)

We want the following probability $P(\hat{p} \leq .82)$

where $p = .84$ $n = 1,000$ so

$$P(\hat{p} \leq .84) = P\left(\frac{\hat{p} - .84}{\sqrt{\frac{.84(1-.84)}{1000}}} \leq \frac{.82 - .84}{\sqrt{\frac{.84(1-.84)}{1000}}}\right) = P(Z \leq 1.73)$$

3. Consider the previous two questions. Suppose now that a sample of 100 subjects are drawn from a *new* population and that 60 of the sampled subjects had an IQs below 116. Give a 95% confidence interval estimate of the true probability of drawing a subject from this population with an IQ below 116. Does this proportion appear to be different than the 84% for the population from questions 1 and 2?

$\hat{p} = .60$ $n = 100$

$$.60 \pm 1.96\sqrt{\frac{.60(1-.60)}{100}} = [.50, .70]$$

Since .84 is not included, this proportion does appear to be different.

4. Let X be binomial with success probability p_1 and n_1 trials and Y be an independent binomial with success probability p_2 and n_2 trials. Let $\hat{p}_1 = X/n_1$ and $\hat{p}_2 = Y/n_2$ be the associated sample proportions. What would be an estimate for the standard error for $\hat{p}_1 - \hat{p}_2$? To have consistent notation with the next problem, label this value $\hat{SE}_{\hat{p}_1 - \hat{p}_2}$.

$$\hat{SE}_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

(adding the two variances and sqr rooting)

5. Consider the previous problem. Suppose that it is known that

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\hat{SE}_{\hat{p}_1 - \hat{p}_2}}$$

is standard normally distributed for large n_1 and n_2. Use this fact to derive a 95% confidence interval for $p_1 - p_2$. (Do not just give an answer, show some steps.)

$$.95 = P\left(-1.96 \le \frac{\hat{p}_1 - \hat{p}_2 - p_1 - p_2}{SE_{\hat{p}_1-\hat{p}_2}} \le 1.96\right)$$

$$= P\left(\hat{p}_1 - \hat{p}_2 + 1.96\, SE_{\hat{p}_1-\hat{p}_2} \ge p_1 - p_2 \ge \hat{p}_1 - \hat{p}_2 - 1.96\, SE_{\hat{p}_1-\hat{p}_2}\right)$$

6. You are in desperate need to simulate standard normal random variables yet do not have a computer available. You do, however, have ten standard six sided dice. Knowing that the mean of a single die roll is 3.5 and the standard deviation is 1.71, describe how you could use the dice to approximately simulate standard normal random variables. (Be precise.)

① Roll the ten dice

② Take the average, say $\bar{X}$

③ Calculate $\dfrac{\bar{X} - 3.5}{1.71/\sqrt{10}} = Z$

This will be approx. normally distributed

The next questions involve the following setting: Suppose that 18 obese subjects were randomized, 9 each, to a new diet pill and a placebo. Subjects' body mass indices (BMIs) were measured at a baseline and again after having received the treatment or placebo for four weeks. The average difference from follow-up to the baseline (followup - baseline) was -3 kg/m^2 for the treated group and 1 kg/m^2 for the placebo group. The corresponding standard deviations of the differences was 1.5 kg/m^2 for the treatment group and 1.8 kg/m^2 for the placebo group.

7. Calculate *and interpret* a 95% confidence interval for the change in BMI for the treated group; assume normality.

Use a one sample t test. Let T be the t quantile for 8 degrees of freedom and 0.975. Then we want:

-3 +/- T 1.5 / sqrt(9) (you solve for the numbers)

For interpretation: we are 95% confident that the mean difference for the treated group is between these two numbers. In this case, the interval was entirely negative, suggesting a weight loss from baseline for the treated group.

8. Does the change in BMI over the two year period appear to differ between the treated and placebo groups? Create the relevant 95% confidence interval *and interpret*. Assume normality and a common variance.

Here we want to do a two sample T test. We are asked to assume a common variance. Therefore, we need to calculate the pooled variance estimate. This is:

Sp = sqrt{(8 * 1 ^ 2 + 8 * 1.5^2) / 16} , which in this case is the sqrt of the average of the two variances. (Note that it is not the average of the standard deviations.) After having calculated Sp, we calculate the interval. Let T be the T .975 quantile for 16 degrees of freedom.

-3 - 1 +/1 T * Sp * sqrt(1 / 9 + 1 / 9) (you solve for the numbers)

Zero again is not in the interval. So the evidence suggest (at 95% confidence) that there was more weight loss for the treated