Chris Piech
CS 109

Section #6
Nov 1, 2021

Section 6: Samples

1. **Warmup**: *populations vs. samples*

   What is the difference between the population variance, σ^2, and sample variance, S^2? What is the difference between sample variance, S^2, and variance of the sample mean, $\text{Var}(\bar{X})$?

2. **Beta Sum**: *beta distribution and sum of RVs*

   What is the distribution of the sum of 100 IID Betas? Let X be the sum

   $$X = \sum_{i=0}^{100} X_i \qquad \text{Where each } X_i \sim \text{Beta}(a = 3, b = 4)$$

   Use theory to give a mathematical result. Note the variance of a Beta:

   $$\text{Var}(X_i) = \frac{ab}{(a+b)^2(a+b+1)} \qquad \text{Where } X_i \sim \text{Beta}(a, b)$$

   What is the probability that the sum is either less than 40 or greater than 44?

3. **Food for thought** *CLT*

   Karel the dog eats an unpredictable amount of food. Every day, the dog is equally likely to eat between a continuous amount in the range 100 to 300 ml. How much Karel eats is independent of all other days. You only have 6.5kg of food for the next 30 days. What is the probability that 6.5kg will be enough for the next 30 days?

4. **Variance of Height among Island Corgis**: *sampling and bootstrapping*

   A colleague has collected samples of heights of corgis that live on two different islands. The colleague collects 50 samples from both islands.

The colleague notes that the sample mean is the same between the two groups: both are around 10 inches. However, island B has a **sample variance** that is 3 in^2 **greater** than island A. The colleague wants to make a scientific claim that corgis on island A have a significantly higher spread of heights than corgis on island B. You are skeptical. It is possible that heights are identically distributed across both islands and that the observed difference in variance was a result of chance and a small sample size, i.e. the **null hypothesis**.

Here is the data. Each number is the height, in inches, of an independently sampled corgi (along with the sample variance):

Island A Corgi Heights ($S^2 = 6.0$):
13, 12, 7, 16, 9, 11, 7, 10, 9, 8, 9, 7, 16, 7, 9, 8, 13, 10, 11, 9, 13, 13, 10, 10, 9, 7, 7, 6, 7, 8, 12, 13, 9, 6, 9, 11, 10, 8, 12, 10, 9, 10, 8, 14, 13, 13, 10, 11, 12, 9

Island B Corgi Heights ($S^2 = 9.1$):
8, 8, 16, 16, 9, 13, 14, 13, 10, 12, 10, 6, 14, 8, 13, 14, 7, 13, 7, 8, 4, 11, 7, 12, 8, 9, 12, 8, 11, 10, 12, 6, 10, 15, 11, 12, 3, 8, 11, 10, 10, 8, 12, 8, 11, 6, 7, 10, 8, 5

a. Warmup: You would like to put error bars around your estimate, based off this sample, that the mean height is 10 inches for Island A Corgis. You use Standard Error of the Mean for error bars. Standard Error makes the size of the error bars equal to the standard deviation of your sample estimate. What is the size of your error bar?
b. Calculate the probability of the null hypothesis, that variance is the same acroos islands, using bootstrapping.

Discuss: How would this calculation be different if you were interested in looking at the statistical significance of the difference in sample mean? 95th percentile?