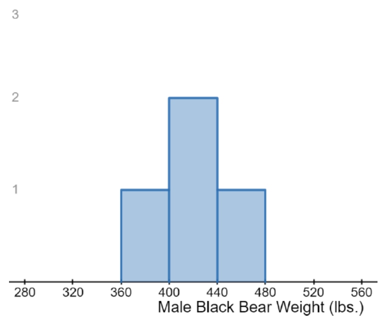
8.2: Distributions of Differences

In this lesson, we will begin to consider statistical methods for comparing independent samples from two populations or experimental treatments. These methods will allow us to compare population means or proportions from two independent groups. To make inferences about differences, we need to use the sampling distribution of differences.

In a previous lesson, we considered the weights of black bears. Black bears weights tend to differ by sex. Female black bears weigh 175 pounds on average, with a standard deviation of 50 pounds, whereas, male black bears weigh 400 pounds on average, with a standard deviation of 40 pounds. Both of the populations are approximately normally distributed.

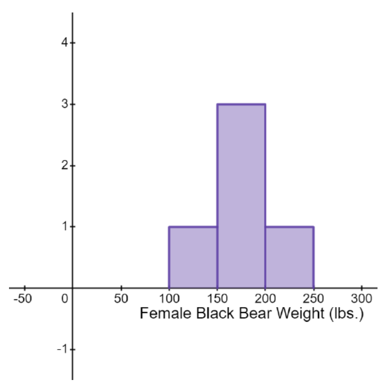
Let's say we are concerned with differences in male and female black bears in a small region in Colorado where the population has dwindled. The population contains only 4 male black bears and 5 female black bears.

Below is a histogram of the 4 male black bear weights. It is very crudely bell-shaped. The mean weight is $\mu_1 = 418.25$ lbs and the standard deviation is $\sigma_1 = 15.943$ lbs. Recall, the variance is the square of the standard deviation. In this case, the variance is $\sigma_1^2 = 254.1875$.



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

Next is a histogram of the 5 female black bear weights. It is also very crudely bell-shaped. The mean weight is $\mu_2 = 167.6$ lbs and the standard deviation is $\sigma_2 = 36.258$ lbs. Recall, the variance is the square of the standard deviation. In this case, the variance is $\sigma_2^2 = 1314.64$.

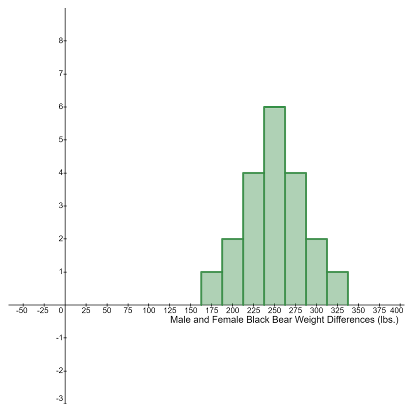


Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

We are interested in the differences in male and female black bear weights for this small population. Let m represent the male black bear weights, and f represent the female black bear weights, then $m - f$ represents the difference in male and female black bear weights. Weight differences are listed for every combination of male and female bear in the table below.

f **values (female black bear weights)**						
		110	**164**	**175**	**165**	**224**
m **values (male black bear weights)**	**418**	308	254	243	253	194
	420	310	256	245	255	196
	395	285	231	220	230	171
	440	330	276	265	275	216

The frequency distribution of weight differences in male and female black bears in the region is given below. It has mean 250.65 lbs, standard deviation 39.608 lbs, and variance 1568.8275.



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

1. Describe the shape of the distribution of differences.

2. Describe the relationship between the mean difference and the means of male and female black bear weights $(\mu_1 \text{ and } \mu_2)$.

3. Describe the relationship between the *variance* of differences and the *variances* of male and female black bear weights $({\sigma_1}^2 \text{ and } {\sigma_2}^2)$.

Summary

- If two independent quantitative variables are approximately normal, then differences between those variables will be approximately normal as well.
- The mean of all differences between those variables is the difference of their means, respectively.
- The variance of all differences between those variables is the sum of their variances, respectively. The standard deviation of differences is the square root of the variance.

Distribution of Differences Between Population Means

- To understand this sampling distribution for the difference in sample means, we just need to think about the sampling distribution for each population. Recall, the Central Limit Theorem states that the sample size must be greater than 30 or the sample must come from a normal population. If we have two samples, then both sample sizes must be greater than 30 or both samples must come from normal populations to ensure that the distribution of differences in sample means is approximately normal.

Consider the sampling distribution of sample means. It has mean $\mu_{\bar{x}} = \mu$, the population mean, and standard error $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, and *variance* ${\sigma_{\bar{x}}}^2 = \left(\frac{\sigma}{\sqrt{n}}\right)^2 = \frac{\sigma^2}{n}$. Extending this for two populations, with population means μ_1 and μ_2, standard deviations σ_1 and σ_2, and sample sizes n_1 and n_2 respectively, we find that

- the mean of the sampling distribution of differences is $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$.
- The sampling distribution of differences has *variance* $\sigma^2_{\bar{x}_1 - \bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$ which implies that the standard error of the sampling distribution of differences is $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

Distribution of Differences Between Population Proportions

- To understand the sampling distribution of the difference in sample proportions, we just need to think about the sampling distribution for each population. Recall, the Central Limit Theorem states that there must be at least 10 expected successes and failures in a sample to ensure its sampling distribution is approximately normal. If we have two samples, then there must be at least 10 expected successes and failures in each sample to ensure that the distribution of differences in sample proportions is approximately normal.

Consider the sampling distribution of sample proportions. It has mean $\mu_{\hat{p}} = p$, the population proportion, standard error $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$, and *variance* $\sigma_{\hat{p}}{}^2 = \left(\sqrt{\frac{p(1-p)}{n}}\right)^2 = \frac{p(1-p)}{n}$.

- Extending this for two populations, with population proportions p_1 and p_2, and sample sizes n_1 and n_2 respectively, we find that the mean of the sampling distribution of differences is $\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2$.
- The sampling distribution of differences has *variance* $\sigma^2_{\hat{p}_1-\hat{p}_2} = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}$ which implies that the standard error of the sampling distribution of differences is $\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$.