Section 9.2

In chapter 9, our goal is to compare two population parameters to each other. We want to know the relationship between the parameters (if they are equal or if one is larger than the other).

Large Independent Samples

Example: An insurance company wants to compare the mean level of current personal liability awards with those from one year earlier. Random samples of cases were selected from each year. The data is summarized below:

Year	Sample Size	Sample Mean	Sample Variance
Current	50	1.32	0.9734
Previous	55	1.04	0.7291

If we want to estimate the true difference between the average amounts of awards over the two years, our best point-estimate of that difference is $\bar{X}_1 - \bar{X}_2$ the difference of the two sample means.

Properties of the Sampling Distribution of $\bar{X}_1 - \bar{X}_2$:

1. $\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$
2. $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
3. **If the sampled populations are normally distributed then so is the distribution of $\bar{X}_1 - \bar{X}_2$ regardless of the sample size.**
4. **If the sampled populations are not normal then we will need to have large sample sizes to ensure that we can approximate the distribution of $\bar{X}_1 - \bar{X}_2$ by the normal distribution.**

Using the properties above and the same structure as we used in section 7.4, we can find a formula for the Confidence Interval for the True Difference Between the Population Means:

(Point Estimator) $\pm$ (Number of Standard Deviations)(Standard Error)

$$\bar{X}_1 - \bar{X}_2 \pm Z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

To use the above formula we must have large sample sizes (n ≥ 30), and the samples must be randomly drawn from independent populations.

Next, we will look at the method of testing hypotheses of the form: $H_0 : (\mu_1 - \mu_2) = D_0$ **vs.** $H_A : (\mu_1 - \mu_2) < D_0$ **(note: the alternative hypothesis may also have > or $\neq$). The D here refers to the specified difference you are looking to detect.**

Many times we want to test that no difference exists. What will the value of D_0 be in those cases? $D_0 = 0$.

The hypotheses will look as described above, and we will have a new **test statistic**, given by:

$$z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - D_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Steps to test a hypothesis:

1. Identify the Null and Alternative hypothesis
2. Calculate the test statistic
3. Determine your rejection region, compute P-value.
4. Make statistical decision
5. Word your final conclusion

Example 1: The manager of a retail clothing store suspects a difference in the mean amount of break time taken by workers during the weekday shifts compared to that of the weekend shifts. It is suspected that the weekday workers take longer breaks on the average. A random sample of 46 weekday workers had a mean $\overline{X}_1 = 53$ minutes of break time per shift and $S_1 = 7.3$ minutes. A random sample of 40 weekend workers had a mean $\overline{X}_2 = 47$ minutes and $S_2 = 9.1$ minutes. Test the manager's suspicion at the 5% level of significance.

Two-Sample Z-Test and CI

```
Sample   N   Mean  StDev  SE Mean
1       46  53.00   7.30      1.1
2       40  47.00   9.10      1.4
```

Difference = μ_1 - μ_2
Estimate for difference: 6.00
95% lower bound for difference: 3.01

Z-Test of difference = 0 (vs >):   Z-Value = 3.34   P-Value = 0.001

Conclusion: At level of significance α = 0.05 we have sufficient evidence to support claim that the weekday workers take longer breaks on the average than people who work on weekend shifts.

Example 2. The local baseball team conducts a study to find the amount spent on refreshments at the ball park. Over the course of the season they gather simple random samples of 50 men and 100 women. For men, the average expenditure was $20, with a standard deviation of $3. For women, it was $15, with a standard deviation of $2. What is the 99% confidence interval for the spending difference between men and women? Assume that the two populations are independent and normally distributed.

Two-Sample Z-Test and CI

Difference = μ_1 - μ_2
Estimate for difference: 5.000
99% CI for difference: (3.759, 6.241)
Z-Test of difference = 0 (vs not =): Z -Value = 10.66   P-Value = 0.000

Answer: The 99% confidence interval is $3.76 to $6.24. That is, we are 99% confident that men outspend women at the ballpark by at least $3.76 and at most $6.24.

Small Independent Samples

What happens when the sample sizes are not greater than or equal to 30? There are two consequences of this:

1. We cannot assume the CLT can give us approximate normality
2. We must know the samples are normally distributed to start with
3. The sample standard deviations may not be reliable estimates of their population counter parts (solution: We will need to use the t-distribution).

In order to use the t-distribution we must assume that the population variances are equal.

Example: Among 28 subjects using the Weight Watchers diet, the mean weight loss after a year was 3.0 pounds with a standard deviation of 4.9 pounds. Among 25 subjects using the Atkins diet, the mean weight loss after one year was 2.1 pounds with a standard deviation of 4.8 pounds. Construct a 95% confidence interval estimate of the difference between the mean weight losses for the two diets (assume weight loss is a normally distributed random variable). Does there appear to be a difference between the effectiveness of the two diets?

To do the above problem using the t-distribution, we must assume that the two variances are equal. It is then reasonable to pool the two sample variances into one sample estimator of σ^2. We call this estimator the pooled sample estimator of σ^2:

$$S_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

To form our confidence interval we will follow the following set of steps:

Step 1 Find $t_{\alpha/2}$ using n_1+n_2-2 as the degrees of freedom $t_{.025,51} = 2.009$

Step 2 Calculate $\bar{X}_1 - \bar{X}_2 = 3.0 - 2.1 = 0.9$

Step 3 Calculate $S_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(28-1)4.9^2 + (25-1)4.8^2}{28+25-2} = 23.6$

Step 4 Form CI: $\bar{X}_1 - \bar{X}_2 \pm t_{\alpha/2}\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$ = **0.9** $\pm\ 2.009\sqrt{\frac{23.6}{28} + \frac{23.6}{25}} = 0.9 \pm 2.69$ CI: **(-1.79, 3.59).**

Answer: Although, the interval suggests that Weight Watchers diet provides larger average weight loss than the Atkins diet, since interval contains 0, we cannot make statistical judgment about a difference between the effectiveness of the two diets. Interval insignificant.

Small-Sample Hypothesis Test

H_0: $\mu_1 - \mu_2 = D_0$

Ha: $\mu_1 - \mu_2 \neq D_0$ *

*Of course, the alternative can be < or >

Test statistic: $t = \frac{(\bar{X}_1 - \bar{X}_2) - D_0}{\sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}}$, *where* $S_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$, and d.f.$= n_1+n_2-2$

Two-Sample T-Test and CI

Sample	N	Mean	StDev	SE Mean
1	28	3.00	4.90	0.93
2	25	2.10	4.80	0.96

Difference = **μ_1 - μ_2**
Estimate for difference: 0.90
95% CI for difference: **(-1.78, 3.58)**
T-Test of difference = 0(vs not =):**T-Value = 0.67 P-Value = 0.503 DF = 51**
Both use Pooled StDev = 4.8532

Example 3. Two growth hormones are being considered. A random sample of 10 rats were given the first hormone and their average weight gain was $\overline{X}_1 = 2.3$ pounds with standard deviation $S_1 = 0.4$ pound. For the second hormone, a random sample of 15 rats showed their average weight gain to be $\overline{X}_2 = 1.9$ pounds with standard deviation $S_2 = 0.2$ pound. Assume the weight gains follow a normal distribution. Using a 10% level of significance, can we say there is a difference in average weight gains for the two growth hormones?

Two-Sample T-Test and CI

Sample	N	Mean	StDev	SE Mean
1	10	2.300	0.400	0.13
2	15	1.900	0.200	0.052

Difference = **μ_1 - μ_2**
Estimate for difference: 0.400
90% CI for difference: (0.156, 0.644)
T-Test of difference = 0 (vs not =): T-Value = 2.93 P-Value = 0.013 DF = 23

Test of Hypothesis: At 10% level of significance we have sufficient evidence to support claim that there is a difference in average weight gains for the two growth hormones.

CI: We are 90% confident that first hormone outperformed the second one in terms of weight gain by at least 0.156 and at most 0.644 pounds.

Example 4. A local bank claims that the waiting time for its customers to be served is the lowest in the area. A competitor's bank checks the waiting times at both banks. The sample statistics are listed below. Test the local bank's claim. Use $\alpha = 0.05$.

Local Bank	**Competitor Bank**
$n_1 = 15$	$n_2 = 16$
$\overline{X}_1 = 5.3$ minutes	$\overline{X}_2 = 5.6$ minutes
$S_1 = 1.1$ minutes	$S_2 = 1.0$ minutes

Two-Sample T-Test and CI

Sample	N	Mean	StDev	SE Mean
1	15	5.30	1.10	0.28
2	16	5.60	1.00	0.25

Difference = **μ_1 - μ_2**
Estimate for difference: -0.300

95% upper bound for difference: 0.344
T-Test of difference = 0 (vs <): T-Value = -0.79 P-Value = 0.217 DF = 29

Conclusion: At 5% level of significance we have insufficient evidence to support local bank claim that the waiting time for its customers to be served is the lowest in the area.

Example 5. Suppose that simple random samples of college freshman are selected from two universities **- 15** students from school A and **20** students from school B. On a standardized test, the sample from school A has an average score of **1000** with a standard deviation of **100**. The sample from school B has an average score of **950** with a standard deviation of **90**. What is the 95% confidence interval for the difference in test scores at the two schools, assuming that test scores came from normal distributions?

Two-Sample T-Test and CI

Difference = $\mu_1 - \mu_2$
Estimate for difference: 50.0
95% CI for difference: (-15.6, 115.6)
T-Test of difference = 0 (vs not =): T-Value = 1.55 P-Value = 0.130 DF = 33
Both use Pooled StDev = 94.3719

We are 95% confident that the true difference in average test scores at the two schools is within: -15.6 to 115.6. Since interval contains 0 the result is statistically insignificant.

Based on P-value = 0.13 we have insufficient evidence to support claim about the difference in average scores for two schools.

Example 6. It is conjectured that classes that use a statistical computer package, such as Minitab, do better in statistics courses than those who don't use technology. A random sample of 24 students uses a Minitab while taking statistics. Another random sample of 28 students taking the same course uses only hand-held calculators. The final average in the course is recorded for each of these students.

Computer

76	93	74	90	90	85	72	81	76	77
69	57	72	100	75	72	86	92	76	79
72	61	80	67						

No computer

74	65	66	85	79	82	80	76	57	75
67	71	55	72	75	89	81	72	88	73
72	64	61	74	63	76	85	72		

Is there sufficient evidence to conclude that students who do not use the computer have lower averages? Use a = .05.

H_0: $\mu_1 - \mu_2 = 0$ **H_a: $\mu_1 - \mu_2 > 0$** **Decision Rule**: Reject H_0 if the calculated **p-value < .05.**

$$\text{Test Statistic: } t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} \ ; \qquad s_p^2 = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}$$

Calculations from Minitab: t = 1.82, p-value = 0.037

Interpretation: At level of significance α = 0.05, we have sufficient evidence that the final average of students using the computer in this statistics course is higher than for those not using the computer.

Section 9.3
Comparing Two Population Means: Matched Pairs

Recall that our goal in the previous section was to be able to detect a difference between two population averages. The example problem below requires the same kind of analysis. We would like to be able to detect if the scores for students taking the FCAT math section improve after completing a series of FCAT prep classes.

Example : Below is a table of FCAT SSS developmental scores for a group of students who were struggling with math in the 3rd grade.

FCAT math scores for 8 students		
Student	After Prep	Before Prep
1	290	275
2	275	270
3	380	370
4	260	245
5	340	325
6	270	260
7	280	270
8	215	200

We want to test the claim: The prep classes work to improve FCAT math SSS scores.

(In symbolic form: $\mu_{After} - \mu_{Before} > 0$)

Using the method from 9.2, we get a test statistic of: t = 0.467, which is not significant. This means we cannot reject the null ($\mu_{After} - \mu_{Before} = 0$), and we conclude that the prep classes are not effective at raising FCAT SSS scores. Does that seem correct when we look at the table of values above? Isn't it true that every student improved their math score after attending the prep classes? Then why did we get this result?

The answer is that the method we used is not valid here. In section 9.2, our assumption was that the two samples were independent, but that is not true here. The two samples above were drawn from the same students. We gave them the FCAT, and then we gave them prep classes and retested the same students.

Okay, so we violated the assumptions—so what? Why does that affect our ability to detect the difference between the two FCAT performances? The answer lies in the quantity:

$$S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = 2{,}581.03$$

This is our pooled variance for the t-test we conducted above. What is this measuring in this case? It is measuring the variation of the FCAT scores between the students—not the difference between the scores each individual earned before and after prep. In other words it is **not** looking at the difference between before and after exams, but rather looking at the differences between student abilities. We know there is a lot of difference between individual student ability, but how is that affecting our hypothesis test?

We are trying to compare the average difference between before and after FCAT scores against the natural variation between FCAT before and after scores. If the distance between the before and after FCAT scores is not much larger than (or is small) compared to the natural variation that occurs between retakes of the FCAT, we will conclude there is no significant improvement due the prep classes.

Consider this simple example: Johnny scores a 170 on his FCAT math section, and Suzy scores a 460 on her FCAT math section. After taking the prep classes, Johnny retakes the FCAT and improves his grade to a 190, while Suzy jumps to a 490.

Suzy's score change = 490 – 460 = +30, Johnny's score change = 190 – 170 = +20
Average score change = + 25

What is the difference in their scores however?

Before difference = Suzy(460) – Johnny(170) = 290
After difference = Suzy(490) – Johnny(190) = 300
Average difference = 295

If you compare these two numbers, it is clear the average score change is quite small compared to the differences between Johnny's and Suzy's FCAT scores, but we do not want to compare these two quantities do we? No, we would want to compare the Average score change against the variation of the individual score changes not the variation of the individual scores.

We need to find a way to ignore these differences between different students' FCAT scores. These differences are not important to us, and we do not want them to obscure the differences between before and after scores that we are interested in. If we can't block out the differences between students, we will never be able to detect the smaller differences that are occurring between before and after test scores. It would be like trying to hear the footsteps of a mouse running across a concert hall floor while a rock concert is being played in the same hall.

Blocking (Matched pairs)

We do have a very simple solution to this problem: we will run a one-sample t-test on the differences between before and after scores: Just subtract each subject's after and before scores
…

Subject	1	2	3	4	5	6	7	8
After	290	275	380	260	340	270	280	215
Before	275	270	370	245	325	260	270	200
Difference	**15**	**5**	**10**	**15**	**15**	**10**	**10**	**15**

Then treat this like a single sample. We can get the average difference $\overline{X_d} = 11.875$, the standard deviation for the differences $S_d = 3.720$, and the number of differences $n_d = 8$. Then we can use the same test statistic for a one-sample t-test:

$$t = \frac{\overline{X_d} - \mu_d}{\frac{S_d}{\sqrt{n_d}}} \quad \text{with degrees of freedom} = n_d - 1$$

Where do we get μ_d? That is the hypothesized value for the true average difference. This leads us to the question, "what will our claims look like?"

We will be conducting our hypothesis test using the following pair of competing claims:

$$H_0 : \mu_d = D_0$$
$$H_A : \mu_d > D_0 \,*$$

*Of course, the alternative can be < or $\neq$

Now let's finish our example properly:

1. Express the original claim symbolically: $\mu_d > 0$ *
2. Identify the Null and Alternative hypothesis: $\begin{matrix} H_0 : \mu_d = 0 \\ H_A : \mu_d > 0 \end{matrix}$
3. Record the data from the problem: $\overline{X_d} = 11.875, S_d = 3.720, n_d = 8, \alpha = 0.05$
4. Calculate the test statistic: $t = \dfrac{\overline{X_d} - \mu_d}{\dfrac{S_d}{\sqrt{n_d}}} = \dfrac{11.875}{\dfrac{3.720}{\sqrt{8}}} \approx 9.029$
5. Determine your rejection region: t > 1.895
6. Statistical decision: Reject the null hypotheses.
7. Word your final conclusion: At level of significance α = 0.05, we have sufficient evidence to support the claim that prep classes are effective at improving student's FCAT math scores.

*Note: our claim that prep classes improve scores indicates that the 'after' exam will be better than the 'before' exam. This means if we form the difference d = After – Before, the differences should be positive, i.e. d > 0.

<u>Confidence Interval for Paired Differences:</u>

$$\left[\overline{X_d} - t_{\alpha/2}\frac{S_d}{\sqrt{n_d}}, \overline{X_d} + t_{\alpha/2}\frac{S_d}{\sqrt{n_d}}\right]$$

Required assumptions:

1. We must assume that the sample was chosen randomly from the target population.
2. We must assume that the population of differences has a normal distribution.

Example 6. A salesman for a shoe company claimed that runners would record quicker times, on the average, with the company's brand of sneaker. A track coach decided to test the claim. The coach selected eight runners. Each runner ran two 100-yard dashes on different days. In one 100-yard dash, the runners wore the sneakers supplied by the school; in the other, they wore the sneakers supplied by the salesman. Each runner was randomly assigned the sneakers to wear for the first run. Use α = 0.05.

Runners	1	2	3	4	5	6	7	8
Company Sneaker	10.8	12.3	10.7	12.0	10.6	11.5	12.1	11.2
School Sneaker	11.4	12.5	10.8	11.7	10.9	11.8	12.2	11.7

Paired T-Test and CI

```
              N      Mean     StDev    SE Mean

Diff.         8   -0.2250    0.2760     0.0976
```

95% CI for mean difference: **(-0.4557, 0.0057)**

T-Test of mean difference = 0 (vs < 0): **T-Value = -2.31; P-Value = 0.027**

Conclusion: At level of significance α = 0.05 we have sufficient evidence to support claim that runners would record quicker times, on the average, with the company's brand of sneaker.

Example 7. Twenty-four males age 25-29 were selected from the Framingham Heart Study. Twelve were smokers and 12 were nonsmokers. The subjects were paired, with one being a smoker and the other a nonsmoker. Otherwise, each pair was similar with regard to age and physical characteristics. Systolic blood pressure readings were as follows:

People	1	2	3	4	5	6	7	8	9	10	11	12
Smokers	122	146	120	114	124	126	118	128	130	134	116	130
Nonsmokers	114	134	114	116	138	110	112	116	132	126	108	116

Use a 5% level of significance to determine whether the data indicate a difference in mean systolic blood pressure levels for the populations from which the two groups were selected. Population of differences is approximately normal.

Paired T-Test and CI

```
               N    Mean   StDev   SE Mean

Diff.          12   6.00    8.40      2.42
```

95% CI for mean difference: **(0.66, 11.34)**

T-Test of mean difference = 0(vs not = 0): **T-Value = 2.47** **P-Value = 0.031**

Conclusion: At level of significance α = 0.05 we have sufficient evidence to support claim that there is a difference in mean systolic blood pressure levels for the two populations.

Example 8. A manufacturer of shock absorbers would like to advertise that their shock absorbers last longer than those produced by its biggest competitor. To see if there is support for such a claim, six of the manufacturer's shocks and six of the competitor's shocks were randomly selected, and one of each brand was installed on the rear wheels of each of six cars. After the cars had been driven 20,000 miles, the strength of each shock absorber was measured. These data are below.

Car	Manufacturer	Competitor
1	8.8	8.4
2	10.5	10.1
3	12.5	12.0
4	9.7	9.3
5	9.6	9.0
6	13.2	13.0

Is there sufficient evidence to conclude that the manufacturer's shocks have a greater mean strength after 20,000 miles of driving than the competitor's? Use a = .01 level of significance.

SOLUTION

The parameter of interest is μ_d, the difference in the mean strength of the manufacturer's shocks and the competitor's shocks after 20,000 miles of driving.

H_0: $\mu_d = 0$ **H_a: $\mu_d > 0$**

Decision Rule: Reject H_0 if the calculated p-value < .01. **Test Statistics**:

$$t = \frac{\overline{X_d} - \mu_d}{\frac{S_d}{\sqrt{n_d}}}$$

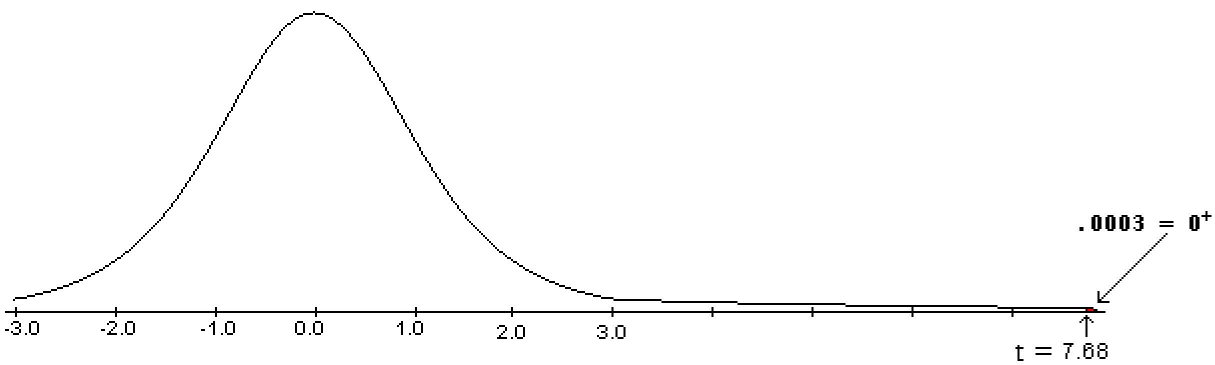
Paired T-Test and CI: Manufacturer, Competitor

```
Paired T for Manufacturer - Competitor

                N      Mean     StDev   SE Mean
Manufacturer    6    10.717     1.752     0.715
Competitor      6    10.300     1.818     0.742
Difference      6    0.4167    0.1329    0.0543

95% lower bound for mean difference: 0.3073
T-Test of mean difference = 0 (vs > 0): T-Value = 7.68  P-Value = 0.000
```

Interpretation: At level of significance α = 0.01, we have sufficient evidence that the manufacturer's shock absorbers have a greater mean strength after being on cars for 20,000 miles than the competitor's shock absorbers.



Section 9.4

Comparing Two Population Proportions Independent Sampling

Example 9: In a randomized controlled trial in Kenya, insecticide treated bed-nets were tested as a way to reduce malaria. Among 343 infants who used the bed-nets, 15 developed malaria. Among 294 infants not using bed-nets, 27 developed malaria Use a 0.01 significance level to test the claim that the incidence of malaria is lower for infants who use the bed-nets. Do the bed-nets seem to work?

To answer this question we need to know what quantities to compare. Let's look at what we have here:

	Bed-Nets	No Nets
X	15	27
n	343	294
$\hat{p}$	0.044	0.092

Clearly, for this sample, bed-nets resulted in a lower infection rate, but is this difference just a coincidence? Maybe this could have happened by chance—after all the number of mosquitoes each group was exposed to was not controlled.

***Sample sizes are large enough when $np \geq 15$ and $nq \geq 15$**

So our **point estimator** will be: $\left(\hat{p}_1 - \hat{p}_2\right)$

The standard error of its sampling distribution will be:

$$\sigma_{(\hat{p}_1-\hat{p}_2)}=\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}\approx\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)},\ where\ \hat{p}=\frac{x_1+x_2}{n_1+n_2}$$

Our competing hypotheses will be: $H_0:(p_1-p_2)=0$ Vs. $H_A:(p_1-p_2)\neq 0$ (<, > are possible also)

Our test statistic will be: $z\approx\dfrac{(\hat{p}_{nets}-\hat{p}_{no-nets})}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}},\ where\ \hat{p}=\dfrac{x_1+x_2}{n_1+n_2}$

Now let us determine if we can support the claim that the bed-nets are effective in preventing malaria infection.

1. Identify the Null and Alternative hypothesis: $H_0:(p_{nets}-p_{no-nets})=0$ $H_A:(p_{nets}-p_{no-nets})<0$

2. Record the data from the problem: $\hat{p}_{\text{nets}}=0.044$, $\hat{p}_{\text{no-nets}}=0.092$, $\alpha=0.01$
3. Calculate the test statistic:

$$z\approx\frac{(\hat{p}_{nets}-\hat{p}_{no-nets})}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}=\frac{-0.048}{\sqrt{0.066(0.934)\left(\frac{1}{343}+\frac{1}{294}\right)}}\approx-2.434$$

4. Determine your rejection region: Z < -2.326
5. Decision: Reject the null hypotheses.
6. **Conclusion: At α = 0.01 we can support the claim that the bed-nets are effective at preventing malaria.**

Confidence Interval for (p_1-p_2):

$$(\hat{p}_1-\hat{p}_2)\pm z_{\alpha/2}\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}\approx(\hat{p}_1-\hat{p}_2)\pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$(0.044-0.092)\pm 2.575\sqrt{\frac{0.044\cdot 0.956}{343}+\frac{0.092\cdot 0.908}{294}}=-0.048\pm 0.001$$

CI: (-0.049, -0.047) We are 99% confident that proportion of infants that was infected while using the bed–nets is about 5% smaller than for those who didn't use the bed-nets.

<u>Example 10</u>. A new insect spray, type A, is to be compared with a spray, type B, that is currently in use. Two rooms of equal size are sprayed with the same amount of spray, one room with A, the other with B. Two hundred insects are released into each room, and after 1 hour the numbers of dead insects are counted." There are 120 dead insects in the room sprayed with A and 90 in the room sprayed with B. Do the data provide enough evidence to indicate that spray A is more effective than spray B? Use α = .05.

Ho: Spray A is not more effective than Spray B. **H_o: $p_A - p_B \le 0$**

Ha: Spray A is more effective than Spray B. **H_a: $p_A - p_B > 0$**

Assumptions: Two independent binomial experiments were performed. Let $\mathbf{P_A}$ be the proportion of insect killed by Spray A and $\mathbf{P_B}$ be the proportion of insect killed by Spray B. Since nA = nB = 200 with $\hat{p}$A =0.6 and $\hat{p}_B$ =0.45, the sampling distribution of $\hat{p}$A - $\hat{p}_B$ is approximately normal.

α = .05, RR: z > 1.645

Test Statistics:

$$z \approx \frac{\left(\hat{p}_{nets} - \hat{p}_{no-nets}\right)}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}, where\ \hat{p} = \frac{x_1 + x_2}{n_1 + n_2} \qquad \hat{p} = \frac{120+90}{200+200} = 0.525$$

$$Z = \frac{0.60-0.45}{\sqrt{0.525\times0.475\left(\frac{1}{200}+\frac{1}{200}\right)}} = 3.038$$

The **p-value** = P(z > 3.04) = .5 - .4988 = .0012. P-value < 0.05

Decision: Reject the null hypothesis.

Conclusion: At level of significance α = .05 there is enough evidence to indicate that

Spray A is more effective than Spray B.

<u>90% confidence interval for $P_A - P_B$</u>

$$\left(\hat{p}_1 - \hat{p}_2\right) \pm z_{\alpha/2}\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}} \approx \left(\hat{p}_1 - \hat{p}_2\right) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1}+\frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$(0.60 - 0.45) \pm 1.645\sqrt{\frac{0.60\times0.40}{200}+\frac{0.45\times0.55}{200}} = 0.15 \pm 0.08 \qquad \mathbf{7\% < P_A - P_B < 23\%}$$

We are 90% confident that Spray A kills between 7% and 23% more insects than Spray B.

Test and CI for Two Proportions

```
Sample   X    N      p̂

1       120   200   0.60
2        90   200   0.45


Difference = PA - PB
Estimate for difference:  0.15
90% lower bound for difference:  0.0867285
Test for difference = 0 (vs > 0):  Z = 3.00   P-Value = 0.001
```