Hypothesis Test for the Difference of Two Independent Population Proportions

Information and questions sourced from:
"Collaborative Statistics" by Dr Barbara Illowsky and Susan Dean.
Available at http://cnx.org/content/col10522/1.40/

And from:
"OpenIntro Statistics - Second Edition" by D Diez, C Barr and M Cetinkaya-Rundel.
Available at https://www.openintro.org/stat/textbook.php?stat_book=os

When comparing Two Independent Population Proportions, we need to note:

1. The two independent samples are simple random samples that are independent.

2. The number of successes is at least five and the number of failures is at least five for each of the samples. (this comes from $np>5$ and $nq>5$, due to using the Normal Approximation to the Binomial distribution)

Comparing two proportions, like comparing two means, is common.
If two estimated proportions are different, it may be due to a difference in the populations or it may be due to chance. A hypothesis test can help determine if a difference in the sample proportions ($p_1 - p_2$) reflects a difference in the population proportions.

The difference of two proportions follows an approximate normal distribution. Generally, the null hypothesis states that the two population proportions are the same. That is, $H_0 : p_1 = p_2$.

To conduct the test of $p_1 = p_2$, we use a pooled proportion, p

This is obtained by pooling all the results together and looking at the total number of successes, divided by the total number of cases. The two examples (overleaf) show this quite clearly, whereas the algebraic formula for p (below) somewhat disguises this fact.

The AH Statistics Formula Booklet cites the following test statistic and formula for p, based upon two samples n_1 and n_2, with sample proportions p_1 and p_2:

z-test for a difference in population proportions:

$$\frac{p_1 - p_2}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim \mathrm{N}(0,1) \text{ where } p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

This document contains 2 examples with full worked solutions, then 5 further questions for you to attempt, each with full worked solutions.

Be sure to note all details that differ when comparing your solutions to those provided.

Example 1

Two types of medication for hives are being tested to determine if there is a difference in the proportions of adult patient reactions. Twenty out of a random sample of 200 adults given medication A still had hives 30 minutes after taking the medication. Twelve out of another random sample of 200 adults given medication B still had hives 30 minutes after taking the medication.
Test at a 1% level of significance.

Overleaf are two solutions:

The first is the **full and comprehensive solution**, derived from 'first principals'

The second is a shortened solution, making use of the formulae in the AH Statistics Booklet

However, we first need to highlight a convention of how to use notation ...

We have previously used the sample mean, $\bar{x}$ as an estimate for the population mean μ
Instead, we could have used $\widehat{\mu}$, where the 'hat' signifies an 'estimate' for μ

We have also used the sample standard deviation, s_{n-1} as an estimate for the population standard deviation, σ
Instead, we could have used $\widehat{\sigma}$, where the 'hat' signifies an 'estimate' for σ

We haven't had to use the hat notation, as the letters x and μ are different, and s and σ are different.

However, with proportions, we should use the hat notation to distinguish between a sample proportion, $\widehat{p}$ which is an estimate for a population proportion, p.

Solution to Example 1 – Full and comprehensive version

Let X_A = **number** of adults still with hives after taking medicine A

So $X_A \sim B(200, p_A)$

And let X_B = **number** of adults still with hives after taking medicine B

So $X_B \sim B(200, p_B)$

$H_0: p_A = p_B$

$H_1: p_A \neq p_B$

Assume H_0 to be true.

α=0.01, two-tailed test.

We approximate X_A with a normal distribution $Y_A \sim N(200p_A, 200p_Aq_A)$ where $q_A = 1 - p_A$

This approximation will be allowed if $200p_A > 5$ and $200q_A > 5$

We define $\frac{Y_A}{200}$ = **proportion** of adults with hives after taking medicine A

So $\frac{Y_A}{200} \sim N\left(p_A, \frac{p_Aq_A}{200}\right)$ and similarly, $\frac{Y_B}{200} \sim N\left(p_B, \frac{p_Bq_B}{200}\right)$

We are interested in the difference between the proportions:

$$\frac{Y_A}{200} - \frac{Y_B}{200} \sim N\left(p_A - p_B, \frac{p_Aq_A}{200} + \frac{p_Bq_B}{200}\right)$$

Now, under H_0, we have assumed that p_A equals p_B, so $p_A - p_B$ is 0.

So, $$\frac{Y_A}{200} - \frac{Y_B}{200} \sim N\left(0, \frac{pq}{200} + \frac{pq}{200}\right)$$ where the pooled sample proportion, p is $\frac{x_A + x_B}{400}$

$$\frac{Y_A}{200} - \frac{Y_B}{200} \sim N\left(0, pq\left(\frac{1}{200} + \frac{1}{200}\right)\right)$$

$$\frac{\frac{Y_A}{200} - \frac{Y_B}{200} - 0}{\sqrt{pq\left(\frac{1}{200} + \frac{1}{200}\right)}} \sim N\left(0, 1^2\right)$$

This gives the test statistic, $z = \dfrac{\hat{p}_A - \hat{p}_B}{\sqrt{pq\left(\frac{1}{200} + \frac{1}{200}\right)}}$

Substituting in the values for $\hat{p}_A = \frac{20}{200}$ and $\hat{p}_B = \frac{12}{200}$ we get $p = \frac{32}{400}$, and

$$z = \frac{\frac{20}{200} - \frac{12}{200}}{\sqrt{\frac{32}{400}\frac{368}{400}\left(\frac{1}{200} + \frac{1}{200}\right)}}$$

$$= 1 \cdot 47442$$

$$p - value = 2 \times P(Z > 1 \cdot 47442)$$

$$= 2 \times 0 \cdot 070184$$

$$= 0 \cdot 140369$$

As the p-value is > 0.01, we do not have evidence to reject H_0 and thus at a 1% level of significance, from the sample data, there is not sufficient evidence to conclude that there is a difference in the proportions of adult patients who did not react after 30 minutes to medication A and medication B.

Solution to Example 1 - Shortened

Let A and B be the subscripts for medication A and medication B.
Then p_A and p_B are the desired population proportions.
We have at least 5 'successes' and 'failures' in each group, so we can proceed with the test.
For a random variable, let $X = P_A - P_B$ = difference in the proportions of adult patients who did not react after 30 minutes to medication A and medication B.

$H_0 : p_A = p_B$
$H_1 : p_A \neq p_B$
Assume H_0 to be true.
α=0.01, two-tailed test.

$$n_A = 200, \quad n_B = 200$$
$$\widehat{p}_A = \frac{20}{200}, \quad \widehat{p}_B = \frac{12}{200}$$

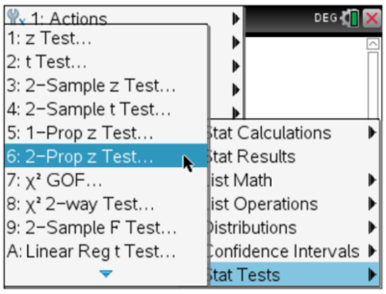
$$\begin{aligned} so\ p &= \frac{n_A\widehat{p}_A + n_B\widehat{p}_B}{n_A + n_B} \\ &= \frac{20+12}{200+200} \\ &= \frac{32}{400} \end{aligned}$$

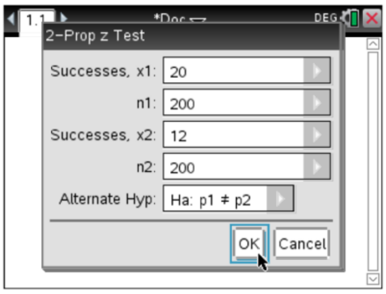
$$\begin{aligned} and\ z &= \frac{\widehat{p}_A - \widehat{p}_B}{\sqrt{pq\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}} \\ &= \frac{\frac{20}{200}-\frac{12}{200}}{\sqrt{\frac{32}{400}\frac{368}{400}\left(\frac{1}{200}+\frac{1}{200}\right)}} \\ &= 1.47442 \end{aligned}$$

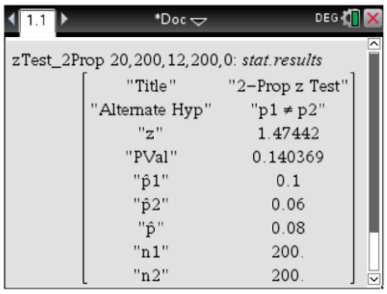
$$\begin{aligned} p-value &= 2 \times P(Z > 1.47442) \\ &= 2 \times 0.070184 \\ &= 0.140369 \end{aligned}$$

As the p-value is > 0.01, we do not have evidence to reject H_0 and thus at a 1% level of significance, from the sample data, there is not sufficient evidence to conclude that there is a difference in the proportions of adult patients who did not react after 30 minutes to medication A and medication B.

Performing this test using a TI-Nspire:







Example 2

In the 2010 Census, 3% of the U.S. population reported being two or more races. However, the percentage varies tremendously from state to state.

(Source: http://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf)

Suppose that two random surveys are conducted. In the first random survey, out of 1000 North Dakotans, only 9 people reported being of two or more races. In the second random survey, out of 500 Nevadans, 17 people reported being of two or more races. Conduct a hypothesis test to determine if the population percentages are the same for the two states or if the percentage for Nevada is statistically higher than for North Dakota.

Solution 2 – Short Version

Let 1 be the subscript for North Dakotans and 2 be the subscript for Nevadans
Then p_1 and p_2 are the desired population proportions.
We have at least 5 'successes' and 'failures' in each group, so we can proceed with the test.

For a random variable, let $X = P_1 - P_2$ = difference in the proportions of adult patients who report being of two or more races.

$H_0 : p_1 = p_2$
$H_1 : p_1 < p_2$ (The words "is statistically higher" tell you the test is one-tailed.)
Assume H_0 to be true.
$\alpha=0.05$, one-tailed test.

$$n_1 = 1000, \quad n_2 = 500$$
$$\widehat{p}_1 = \frac{9}{1000}, \quad \widehat{p}_2 = \frac{17}{500}$$

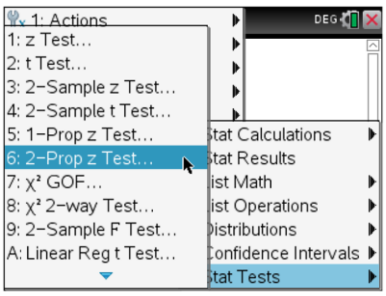
$$so\ p = \frac{n_1\widehat{p}_1 + n_2\widehat{p}_2}{n_1 + n_2} = \frac{9+17}{1000+500} = \frac{26}{1500}$$

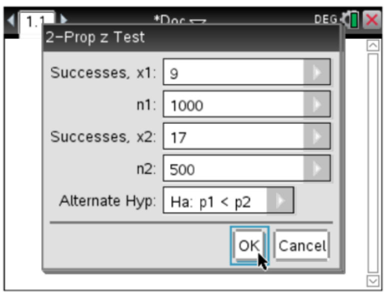
$$and\ z = \frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\frac{9}{1000} - \frac{17}{500}}{\sqrt{\frac{26}{1500}\frac{1474}{1500}\left(\frac{1}{1000} + \frac{1}{500}\right)}} = -3.49732$$

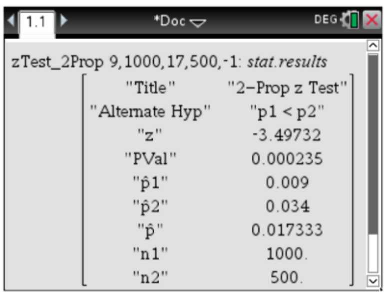
$$p - value = P(Z < -3.49732) = 0.000235$$

As the p-value is < 0.05, we have evidence to reject H_0 and thus at a 5% level of significance, from the sample data, the percentage of Nevadans who claim to be of two or more races is higher than that of North Dakotans.

Performing this test using a TI-Nspire:







Question 1

We have data on sleep deprivation rates of Californians and Oregonians. The proportion of California residents who reported insufficient rest or sleep during each of the preceding 30 days is 8.0%, while this proportion is 8.8% for Oregon residents. These data are based on simple random samples of 11,545 California and 4,691 Oregon residents.
Conduct a hypothesis test to determine if these data provide strong evidence the rate of sleep deprivation is different for the two states.

Solution for Question 1 – Short Version

Let C be the subscript for Californians and O be the subscript for Oregonians
Then p_C and p_O are the desired population proportions.
We have at least 5 ‘successes’ and ‘failures’ in each group, so we can proceed with the test.

For a random variable, let X = $P_C - P_O$ = difference in the proportions of adult patients who report being sleep deprived

$H_0 : p_C = p_O$
$H_1 : p_C \neq p_O$
Assume H_0 to be true.
α=0.05, two-tailed test.

$$n_C = 11545, \quad n_O = 4691$$
$$\widehat{p}_C = 0.08, \quad \widehat{p}_O = 0.088$$

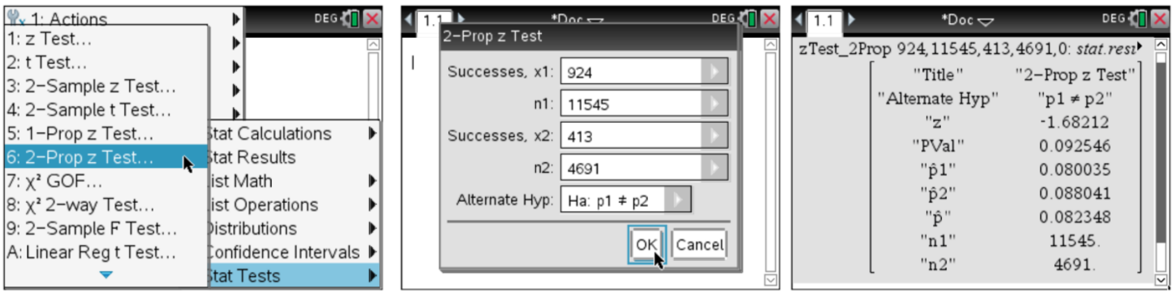
$$\begin{aligned} so\ p &= \frac{n_C\widehat{p}_C + n_O\widehat{p}_O}{n_C + n_O} \\ &= \frac{11545 \times 0.08 + 4691 \times 0.088}{11545 + 4691} \\ &= 0.082311 \end{aligned}$$

$$\begin{aligned} and\ z &= \frac{\widehat{p}_C - \widehat{p}_O}{\sqrt{pq\left(\frac{1}{n_C} + \frac{1}{n_O}\right)}} \\ &= \frac{0.08 - 0.088}{\sqrt{0.082311 \times 0.917689\left(\frac{1}{11545} + \frac{1}{4691}\right)}} \\ &= -1.68113 \end{aligned}$$

$$\begin{aligned} p - value &= 2 \times P(Z < -1.68113) \\ &= 2 \times 0.046368 \\ &= 0.092737 \\ &> 0.05 \end{aligned}$$

As the p-value is > 0.05, we do not have evidence to reject H_0 and so we conclude that the proportion of sleep deprived Californians is the same as those in Oregon.

Performing this test using a TI-Nspire:



Note the requirement to round 0.08×11545=923.6≈924 so that an integer is entered for the number of success, and similarly for 0.088×4691=412.808≈413.

Question 2

A 2010 survey asked 827 randomly sampled registered voters in California "Do you support? Or do you oppose? Drilling for oil and natural gas off the Coast of California? Or do you not know enough to say?"
Below is the distribution of responses, separated based on whether or not the respondent graduated from college. *(Survey USA, Election Poll #16804, data collected July 8-11, 2010.)*

	College Grad	
	Yes	No
Support	154	132
Oppose	180	126
Do not know	104	131
Total	438	389

(a) What percentage of college graduates and what percentage of the non-college graduates in this sample do not know enough to have an opinion on drilling for oil and natural gas off the Coast of California?

(b) Conduct a hypothesis test to determine if the data provide strong evidence that the proportion of college graduates who do not have an opinion on this issue is different than that of non-college graduates.

(c) Conduct a hypothesis test to determine if the data provide strong evidence that the proportion of college graduates who support offshore drilling in California is different than that of non-college graduates.

Solution for Question 2 – Short Version

a) College graduates who 'do not know' = 104/438 = 23.7%
Non- College graduates who 'do not know' = 131/389= 33.7%

b) Let C be the subscript for College Graduates and N be the subscript for Non-College Graduates. Then p_C and p_N are the desired population proportions.
We have at least 5 'successes' and 'failures' in each group, so we can proceed with the test.

For a random variable, let $X = P_C - P_N$ = difference in the proportions of voters who 'do not know'
$H_0 : p_C = p_N$
$H_1 : p_C \neq p_N$
Assume H_0 to be true. α=0.05, two-tailed test.

$$n_C = 438, \quad n_N = 389$$
$$\widehat{p}_C = \frac{104}{438}, \quad \widehat{p}_N = \frac{131}{389}$$

$$\begin{aligned} so\ p &= \frac{n_C \widehat{p}_C + n_N \widehat{p}_N}{n_C + n_N} \\ &= \frac{104+131}{438+389} \\ &= \frac{235}{827} \end{aligned}$$

$$\begin{aligned} and\ z &= \frac{\widehat{p}_C - \widehat{p}_N}{\sqrt{pq\left(\frac{1}{n_C}+\frac{1}{n_N}\right)}} \\ &= \frac{\frac{104}{438}-\frac{131}{389}}{\sqrt{\frac{235}{827}\times\frac{592}{827}\left(\frac{1}{438}+\frac{1}{389}\right)}} \\ &= -3.16081 \end{aligned}$$

$$\begin{aligned} p-value &= 2\times P(Z<-3.16081) \\ &= 2\times 0.000787 \\ &= 0.001573 \\ &< 0.05 \end{aligned}$$

As the p-value is less than 0.05, we have evidence to reject H_0 and conclude that the proportions of 'do not know' are different between the two populations of voters.

/cont.

c) For a random variable, let X = $P_C - P_N$ = difference in the proportions of voters who support offshore drilling.
We have at least 5 'successes' and 'failures' in each group, so we can proceed with the test.

$H_0 : p_C = p_N$
$H_1 : p_C \neq p_N$
Assume H_0 to be true. α=0.05, two-tailed test.

$$n_C = 438, \quad n_N = 389$$
$$\widehat{p}_C = \frac{154}{438}, \ \widehat{p}_N = \frac{132}{389}$$

$$so\ p = \frac{n_C\widehat{p}_C + n_N\widehat{p}_N}{n_C + n_N} = \frac{154+132}{438+389} = \frac{286}{827}$$

$$and\ z = \frac{\widehat{p}_C - \widehat{p}_N}{\sqrt{pq\left(\frac{1}{n_C}+\frac{1}{n_N}\right)}} = \frac{\frac{154}{438}-\frac{132}{389}}{\sqrt{\frac{286}{827}\times\frac{541}{827}\left(\frac{1}{438}+\frac{1}{389}\right)}} = 0.370174$$

$$\begin{aligned} p-value &= 2\times P(Z > 0.370174) \\ &= 2\times 0.355627 \\ &= 0.711253 \\ &> 0.05 \end{aligned}$$

As the p-value is > 0.05, we cannot reject the null hypothesis and conclude that the proportions of those who support offshore drilling is the same in both groups of voters.

Question 3

The National Sleep Foundation conducted a survey on the sleep habits of randomly sampled transportation workers and a control sample of non-transportation workers. The results of the survey are shown below.(National Sleep Foundation, 2012 Sleep in America Poll: Transportation Workers Sleep, 2012.)

		Transportation Professionals			
	Control	Pilots	Truck Drivers	Train Operators	Bux/Taxi/Limo Drivers
Less than 6 hours of sleep	35	19	35	29	21
6 to 8 hours of sleep	193	132	117	119	131
More than 8 hours	64	51	51	32	58
Total	292	202	203	180	210

Conduct a hypothesis test to evaluate if these data provide evidence of a difference between the proportions of truck drivers and non-transportation workers (the control group) who get less than 6 hours of sleep per day, i.e. are considered sleep deprived.

For this question, construct a **full and comprehensive solution**, as detailed previously in Example 1. How much can you do <u>before</u> looking back at example 1?!

Solution to Question 3 – Full and Comprehensive Solution

Let X_T = **number** of truck drivers who get less than 6 hours sleep
So $X_T \sim \mathrm{B}(203, p_T)$
And let X_C = **number** of non-transportation (control) drivers who get less than 6 hours sleep
So $X_C \sim \mathrm{B}(292, p_C)$

$\mathrm{H}_0 : p_T = p_C$
$\mathrm{H}_1 : p_T \neq p_C$
Assume H_0 to be true.
α=0.05, two-tailed test.
We approximate X_T with a normal distribution $Y_T \sim \mathrm{N}(203p_T, 200p_Tq_T)$ where $q_T = 1-p_T$
This approximation is allowed as $203p_T>5$ and $203q_T>5$

We define $\frac{Y_T}{203}$ = **proportion** of truck drivers who get less than 6 hours sleep

So $\frac{Y_T}{203} \sim N\left(p_T, \frac{p_Tq_T}{203}\right)$ and similarly, $\frac{Y_C}{292} \sim N\left(p_C, \frac{p_Cq_C}{292}\right)$

We are interested in the difference between the proportions:

$$\frac{Y_T}{203} - \frac{Y_C}{292} \sim N\left(p_T - p_C, \frac{p_Tq_T}{203} + \frac{p_Cq_C}{292}\right)$$

Now, under H_0, we have assumed that p_T equals p_C, so $p_T - p_C$ is 0.

So, $\frac{Y_T}{203} - \frac{Y_C}{292} \sim N\left(0, \frac{pq}{203} + \frac{pq}{292}\right)$ where the pooled sample proportion, p is $\frac{x_T + x_C}{495}$

$$\frac{Y_T}{203} - \frac{Y_C}{292} \sim N\left(0, pq\left(\frac{1}{203} + \frac{1}{292}\right)\right)$$

$$\frac{\frac{Y_T}{203} - \frac{Y_C}{292} - 0}{\sqrt{pq\left(\frac{1}{203} + \frac{1}{292}\right)}} \sim N\left(0, 1^2\right)$$

This gives the test statistic, $z = \frac{\hat{p}_T - \hat{p}_C}{\sqrt{pq\left(\frac{1}{203} + \frac{1}{292}\right)}}$

Substituting in the values for $\hat{p}_T = \frac{35}{203}$ and $\hat{p}_C = \frac{35}{292}$ we get $p = \frac{70}{495}$, and

$$z = \frac{\frac{35}{203} - \frac{35}{292}}{\sqrt{\frac{70}{495}\frac{425}{495}\left(\frac{1}{203} + \frac{1}{292}\right)}}$$
$$= 1 \cdot 65036$$

$$p - value = 2 \times P(Z > 1 \cdot 65036)$$
$$= 2 \times 0 \cdot 049435$$
$$= 0 \cdot 09887$$

As the p-value is > 0.05, we do not have evidence to reject H_0 and thus at a 5% level of significance, from the sample data, there is not sufficient evidence to conclude that truck drivers and non-transportation drivers have different proportions of sleep deprivation.

Question 4

A “social experiment” conducted by a TV program questioned what people do when they see a very obviously bruised woman getting picked on by her boyfriend. On two different occasions at the same restaurant, the same couple was depicted. In one scenario the woman was dressed “provocatively” and in the other scenario the woman was dressed “conservatively”. The table below shows how many restaurant diners were present under each scenario, and whether or not they intervened.

		Scenario		
		Provocative	Conservative	Total
Intervene	Yes	5	15	20
	No	15	10	25
	Total	20	25	45

Explain why the sampling distribution of the difference between the proportions of interventions under provocative and conservative scenarios does not follow an approximately normal distribution.

Solution to Question 4

This is not a randomized experiment, and it is unclear whether people would be affected by the behaviour of their peers. That is, independence may not hold.

Additionally, there are only 5 interventions under the provocative scenario, so we are on the 'borderline' of being able to progress with confidence that the Normal approximation to the underlying Binomial distribution is a good one.

Question 5

The following question is based upon data provided by S. Lockman et al. "Response to antiretroviral therapy after a single, peripartum dose of nevirapine". In: Obstetrical & gynecological survey 62.6 (2007), p. 361.

In July 2008 the US National Institutes of Health announced that it was stopping a clinical study early because of unexpected results. The study population consisted of HIV-infected women in sub-Saharan Africa who had been given single dose Nevaripine (a treatment for HIV) while giving birth, to prevent transmission of HIV to the infant. The study was a randomized comparison of continued treatment of a woman (after successful childbirth) with Nevaripine vs. Lopinavir, a second drug used to treat HIV.
240 women participated in the study; 120 were randomized to each of the two treatments. Twenty-four weeks after starting the study treatment, each woman was tested to determine if the HIV infection was becoming worse (an outcome called virologic failure). Twenty-six of the 120 women treated with Nevaripine experienced virologic failure, while 10 of the 120 women treated with the other drug experienced virologic failure.

(a) Create a two-way table presenting the results of this study.

(b) State appropriate hypotheses to test for independence of treatment and virologic failure.

(c) Conduct the hypothesis test and state an appropriate conclusion. Be sure to clearly state that you have verified the necessary conditions for the test.

Solution to Question 5 – Shortened version

a) Two way table:

		Virol. failure		
		Yes	No	Total
Treatment	Nevaripine	26	94	120
	Lopinavir	10	110	120
	Total	36	204	240

b) If the treatments are independent, then they should have different levels of viral failure
Let N be the subscript for Nevaripine and L be the subscript for Lopinavir.
Then p_N and p_L are the desired population proportions.
And so the hypotheses are:
$H_0 : p_N = p_L$
$H_1 : p_N \neq p_L$

c) Conditions Checking:

- We have at least 5 'successes' and 'failures' in each group, so we can proceed with the test.
- Also, we had random assignment of patients to each group, so the observations in each group are independent.
- If the patients in the study are representative of those in the general population (something that is impossible to check with the given information), then we can also confidently generalise the findings of this hypothesis to the general population.

For a random variable, let $X = P_N - P_L$ = difference in the proportions of viral failures
$H_0 : p_N = p_L$
$H_1 : p_N \neq p_L$
Assume H_0 to be true. $\alpha=0.05$, two-tailed test.

$$n_N = 120, \quad n_L = 120$$

$$\widehat{p}_N = \frac{26}{120}, \quad \widehat{p}_L = \frac{10}{120}$$

$$so\ p = \frac{n_N\widehat{p}_N + n_L\widehat{p}_L}{n_N + n_L} = \frac{26+10}{120+120} = \frac{36}{240}$$

$$and\ z = \frac{\widehat{p}_N - \widehat{p}_L}{\sqrt{pq\left(\frac{1}{n_N}+\frac{1}{n_L}\right)}} = \frac{\frac{26}{120}-\frac{10}{120}}{\sqrt{\frac{36}{240}\times\frac{204}{240}\left(\frac{1}{120}+\frac{1}{120}\right)}} = 2.89241$$

$$p-value = 2\times P(Z > 2.89241) = 2\times 0.001912 = 0.003823 < 0.05$$

As the p-value is less than 0.05, we have evidence to reject H_0 and conclude that the proportions of viral failure are different for the two treatments.

Important Feature to Notice:
You may have noticed that this two-tailed Hypothesis test for proportions looks rather similar to conducting a Chi-Squared test for independence in a Contingency Table. And you'd be right – they are, in effect, the same test. Look at the p-value from the screenshot on the right, and compare to the above solution. But a one-tailed hypothesis test of proportions would not be the same as a Chi-Squared test. You should be able to see why for yourself.