

A Powerful Problem

Consider the scenario in which a cereal company claims that 20% of all its cereal boxes contain a voucher for a free DVD rental. A group of students believes the company is cheating and the proportion of all boxes with the vouchers is less than 0.20. They decide to collect some data to perform a test of significance with the following hypotheses.

$$H_0 : p = 0.20$$

$$H_A : p < 0.20$$

where p = the proportion of all boxes with the voucher

They collect a random sample of 65 boxes and find 11 boxes with the voucher. Using a One Proportion z -test, the students calculate a p -value = 0.27 and conclude that they do not have enough evidence to say that the proportion of all boxes is less than 0.20. Although the company may be cheating its customers, the students do not have convincing evidence that this is the case.

```
1-PropZTest
PROP<.2
z=-.6201736729
P=.2675716664
P=.1692307692
n=65
```

PART I: WILL THE STUDENTS UNCOVER CORPORATE WRONGDOING?

The question this handout addresses is the following.

If the company is in fact cheating its customers,
how likely would it be for a test based on 65 boxes to catch the company?

1. Suppose the students used a significance level of $\alpha = .05$ in conducting their test. Explain what this significance level represents and how it affects the decision they made.

2. The students found 11 out of 65 boxes with vouchers and did not conclude the company was cheating. How many boxes with vouchers out of 65 would they have needed to find in order to conclude that the company is cheating? Use trial and error with One Proportion z -test on your calculator to find the range of number of voucher boxes that would lead to a conclusion of corporate wrongdoing.

The assumption in this handout is the company is cheating, and the question is how likely would it be for the students' 65 box test with $\alpha = .05$ to detect this cheating.

A natural question to ask then is: How badly is the company cheating? Pretend the company's proportion of all boxes with vouchers is really 0.15 ($p = 0.15$). If a 65 box test using $\alpha = .05$ were performed, would the students correctly conclude that the company is cheating ($p < 0.20$)? Let's find out.

- You will sample 65 cereal boxes from a population in which 15% of all boxes contain a voucher. The calculator command below, which can be found under MATH – PRB, simulates random sampling from this population. Run this command to take a sample of 65.

`randBin(65, .15)`

In question #2, you should have arrived at the following rule for concluding that the company is cheating. (Recall this rule is based on the significance level of $\alpha = .05$.)

Conclude the company is cheating if you obtain
7 or fewer boxes with vouchers out of 65.

- In your 65 box trial from question #3, how many boxes with vouchers did you obtain? Based on your result, did you have enough evidence to conclude that the company is cheating?
- Repeat your simulation from question #3 twenty times and record your results in the table below. (To repeat the command `randBin(65,.15)` simply press ENTER.)

Trial	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Voucher Boxes																				

- Remember, the assumption in the simulation is the company is cheating— $p = 0.15$. Out of your 20 trials in question #5, in how many of them did you conclude that the company is cheating?

The proportion of 20 trials in which you concluded the company is cheating is your estimate of the probability of the 65 box test with $\alpha = .05$ detecting the company's placement of vouchers in only 15% of its boxes.

7. Combine your results as a class in order to make an estimate based on more trials. What is the class estimate of the probability of the 65 box test with $\alpha = .05$ detecting the company's placement of vouchers in only 15% of its boxes?

The probability you just calculated is an estimate of the POWER of a One Proportion z-test with $n = 65$ and $\alpha = .05$ against the alternative of $p = 0.15$.

8. Using your result in question #7, comment on the students' ability to detect a company that puts vouchers in only 15% of its boxes by using a 65 box test with $\alpha = .05$.

1000 trials of the simulation from question #3 were conducted using computer software. In 226 of these trials, 7 or fewer boxes with vouchers were found, and thus in 22.6% of the trials it was concluded the company was cheating. So, it is not all that likely for the students' 65 box test using $\alpha = .05$ to detect a company whose proportion of all boxes with vouchers is 0.15!

You have seen what power represents in this scenario. A more general definition of POWER is given below and can be applied to any situation in which a test of significance is performed.

The POWER of a test of significance against a given alternative is the probability that it rejects the null hypothesis.

PART II: WHAT IF THEY CHANGED THE SAMPLE SIZE?

The students randomly selected 65 boxes in performing their test of significance. It was calculated earlier, via simulation, that the students' test, using $\alpha = .05$, has a power of approximately 0.226 against the alternative hypothesis of $p = 0.15$.

What would happen to the power against $p = 0.15$ if the sample size was increased? This is the subject of the investigation below.

Suppose the students decide to perform a second test, only this time they will randomly select 130 boxes. If the students use the same hypotheses as in their first 65 box test and use $\alpha = .05$, they would have the following rule for concluding the company is cheating.

Conclude the company is cheating if you obtain 18 or fewer boxes with vouchers out of 130.

- 9. Verify that the rule given above for concluding the company is cheating is correct.
- 10. Pretend the company is cheating with $p = 0.15$. Simulate the selection of a random sample of 130 cereal boxes from a population in which 15% of all boxes contain a voucher. How many boxes with vouchers did you obtain? Based on your result, do you have enough evidence to conclude that the company is cheating?
- 11. Repeat your simulation from question #10 twenty times and record your results in the table below.

Trial	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Voucher Boxes																				

- 12. Remember, the assumption in the simulation is the company is cheating— $p = 0.15$. Out of your 20 trials in question #11, in how many of them did you conclude that the company is cheating?
- 13. Comment on which sample size— $n = 65$ or $n = 130$ —would result in the higher power against the alternative $p = 0.15$.

PART III: WHAT IF THE VALUE OF H_A CHANGES?

What if the company is cheating with a value of p different from 0.15? Good question. Suppose the company is cheating with a value of $p = 0.10$ —that is, only 10% of all boxes have the vouchers inside. How likely will it be for the students' 65 box test to catch the company?

14. Predict how the power of the 65 box test using $\alpha = .05$ against the alternative $p = 0.10$ will compare to its power against the alternative $p = 0.15$.

15. In order to test your prediction, run 20 trials of the simulation from question #3—be sure to change the probability used from .15 to .10. Record your results in the table below.

Trial	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Voucher Boxes																				

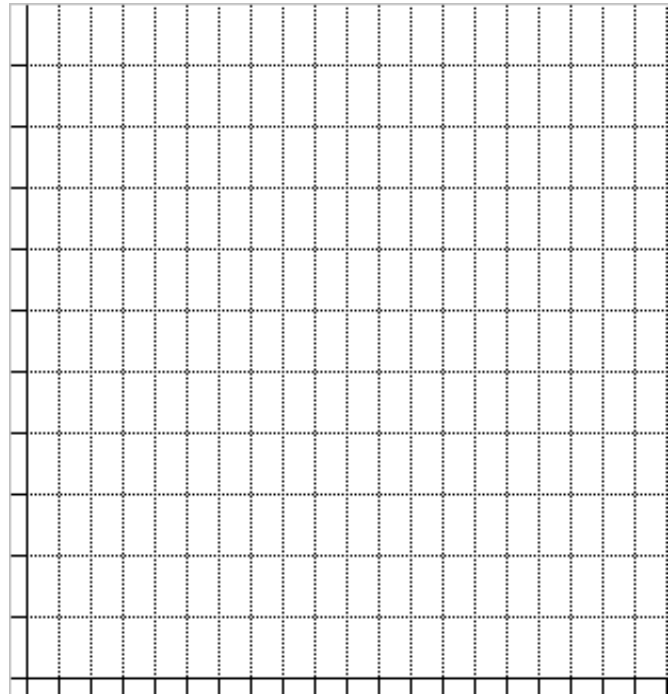
16. Remember, the assumption in the simulation is the company is cheating— $p = 0.10$. Out of your 20 trials in question #15, in how many of them did you conclude that the company is cheating? (Recall how many boxes with vouchers the 65 box test must find in order to conclude the company is cheating.)

17. Comment on which alternative— $p = 0.10$ or $p = 0.15$ —the 65 box test with $\alpha = .05$ has a higher power against and how the concept of power agrees with your intuitive sense about which value of p would be easier to detect.

The alternative hypothesis specifies any value less than 0.20. The table below lists possible alternative values, along with an estimate of the power of the 65 box test with $\alpha = .05$ against each alternative. These estimates of the power were obtained by simulating 1000 trials of 65 box tests (using $\alpha = .05$) for each alternative.

$H_A: p = \text{---}$	Power against H_A (estimated)
.19	.066
.18	.080
.17	.129
.16	.150
.15	.226
.14	.292
.13	.358
.12	.464
.11	.579
.10	.673
.09	.775
.08	.846
.07	.921
.06	.963
.05	.982
.04	.994
.03	1
.02	1
.01	1
0	1

18. On the axes below, construct a scatter plot of the data in the table at left. Put p on the x -axis and Power on the y -axis.



19. Comment on how the distance between $p = 0.2$ and the alternative values of p affects the power of the 65 box test with $\alpha = .05$. Specifically, as the distance increases, how does the power of the test change?

PART IV: HOW DOES SIGNIFICANCE LEVEL AFFECT POWER?

Statisticians are interested in the power of their tests of significance. Knowing how much power a test has against a certain alternative gives them an idea of how likely it is for their significance test to reject the null hypothesis correctly if a certain alternative is true. You have seen that the sample size and the distance between the hypothesized value and the alternative value of p affect the power of a test. There is one more factor that affects the power.

It was assumed the students used a significance level of $\alpha = .05$ in performing their 65 box test. But what if they used a significance level of $\alpha = .01$? $\alpha = .10$? You will investigate below how changing the significance level affects power in this section.

20. Earlier, you discovered that in a 65 box test using $\alpha = .05$, the students would conclude the company was cheating if they obtained 7 or fewer boxes with vouchers. Use trial and error with One Proportion z -test on your calculator to find the upper limit of voucher boxes students would need to find in order to conclude the company is cheating for the other significance levels. Fill in the table.

Significance Level (α)	.01	.05	.10
Upper Limit of voucher boxes in order to conclude cheating		7	

Pretend the company is cheating—they are putting the voucher in only 10% of all boxes. Under this assumption, 65 boxes from the population were randomly sampled in 1000 separate instances. In each of the 1000 trials, the number of voucher boxes obtained was recorded. The results from the 1000 random samples taken from the 10% voucher population are recorded in the table below.

Number of Voucher Boxes	Frequency
0	1
1	6
2	20
3	68
4	106
5	162
6	173
7	160
8	114
9	77
10	64
11	26
12	13
13	8
14	1
15	0
16	1

21. Based on the results of the 1000 random samples, calculate estimates of the power of the 65 box test against an alternative of $p = 0.10$ for the different significance levels. Fill in the table.

Significance Level (α)	.01	.05	.10
Estimate of POWER against $p = 0.10$			

22. Comment on how the significance level of a test of significance affects the power of the 65 box test against an alternative of $p = 0.10$. Specifically, as the significance level increases, how does the power change?

PART V: THE THREE FACTORS AFFECTING POWER

As you have seen, there are three factors that affect the power of a test of significance. They are

- I. The sample size (n).
- II. The true value of the population characteristic of interest.
- III. The significance level (α).

23. To summarize the effect these three factors have on power, fill in the table below.

	What happens to the Power?
When the sample size increases...	
When the distance between the hypothesized and alternative values of p increases...	
When the significance level increases...	

In general, statisticians determine what alternative value it is important for them to detect, and select a sample size for their study that gives them the power they desire against that alternative. Thus, a common way for a statistician to adjust the power against a particular alternative is to adjust the sample size.

PART VI: ADDITIONAL TERMINOLOGY

You have seen that power of a test of significance against a given alternative value is the probability that the test rejects the null hypothesis. In the realm of statistical decision-making, there are two other common terms used that are defined below.

A **TYPE I ERROR** occurs if the null hypothesis is rejected when the null hypothesis is true.

24. Explain what a Type I error is in the context of the problem in this handout.

25. The probability a test of significance will lead to a Type I error is equal to the significance level (α) of the test. Explain why this relationship is true.

A **TYPE II ERROR** occurs if the null hypothesis is NOT rejected when the null hypothesis is false.

26. Explain what a Type II error is in the context of the problem in this handout.

27. The probability a test of significance will lead to a Type II error is denoted by the Greek letter β . Explain what the relationship is between β and Power.