AP STATS

by Dr. Bob Stephenson
and Dr. Hal Stern

1 Question

How important is randomization?

2 Activity

Across the midwest farmers are constantly looking for that competitive edge that will increase profits. Responding to this demand, seed companies have developed, through cross breeding, hybrid varieties of corn with higher and higher yields. More recently, through genetic engineering, there are now corn varieties that are resistant to the affects of herbicide residue and others that can combat pests like the European corn borer. Once a variety of corn is developed, the true test of its values comes in field trials. A field trial is a designed experiment used to compare varieties of corn (or soybeans, or wheat, etc.) in terms of average yield (or some other measure of quality). Sir Ronald Fisher developed many of the methods of applied statistics while analyzing agricultural field experiments at Rothamsted in England. The following activity simulates an agricultural field experiment, or field trial, conducted to compare two varieties of corn, A and B.

Class Activity:

Researchers at a large seed company are planning a field trial to compare two hybrid varieties of corn. The response of interest is the yield, in bushels per acre. The better variety will be the one with the highest yields but the researchers recognize that variation in soil composition, fertility and drainage will have effects on the growth of plants and thus yield. There is a filed with 36 plots available for the experiment. On 18 plots variety A will be planted and on the other 18 plots variety B will be planted. The researchers wish to see if the two varieties have equal yields, on average, or if the two varieties differ significantly. If the two varieties really do differ, the researchers would like their experiment and the subsequent statistical analysis to detect this true difference. The ability of a statistical procedure to detect a true difference is called the power of the procedure. The researchers must decide how to assign the varieties to the plots.

- **Convenience Assignment**

  It is easiest to plant one variety on 18 plots on one side of the field and the other variety on the 18 plots on the other side. Modern machinery cam plant up to 18 rows at a time, so planting in this way can be done in one or two passes through the field. Below is a picture of such and assignment and the yields, in bushels per acre, for each plot.

A 130	A 149	A 139	B 155	B 137	B 145
A 149	A 133	A 152	B 131	B 147	B 136
A 141	A 156	A 137	B 146	B 132	B 148
A 150	A 142	A 155	B 136	B 152	B 133
A 139	A 155	A 139	B 147	B 137	B 153
A 155	A 138	A 150	B 137	B 145	B 136

Summary	n	mean	std. dev
A	18	144.9	8.29
B	18	141.8	7.65

Based on this assignment, by convenience, does one variety appear to have a larger mean yield? Is there a significant difference in mean yields between the two corn varieties?

- **Systematic Assignment**

Many people think that an alternating sequence is a random, or at least an unbiased, sequence. For example, when assigning participants to treatment and control, taking every other participant (alternating) for the treatment group appears random. However, if participants are lined up alternating between male and female then all the males will be in one group and all the females in the other. Gender and group would be completely confounded. That is the effects of treatment and control are inseparable from gender effects. In a field, an alternating pattern would be like a checkerboard. Below is a picture of such an alternating pattern and the yields, in bushels per acre, for each plot.

A 130	B 137	A 139	B 155	A 149	B 145
B 137	A 133	B 140	A 143	B 147	A 148
A 141	B 144	A 137	B 146	A 144	B 148
B 138	A 142	B 143	A 148	B 152	A 145
A 139	B 143	A 139	B 147	A 149	B 153
B 143	A 138	B 138	A 149	B 145	A 148

Summary	n	mean	std. dev
A	18	142.3	5.75
B	18	144.5	5.37

Based on this assignment, alternating, does one variety appear to have a larger mean yield? Is there a significant difference in mean yields between the two corn varieties?

Discuss the results from the analysis of the convenience assignment data and those from the analysis of the alternating assignment data. Some may find it a bit disturbing that B appears better for one assignment while A appears better for the other. Of course, this could be due to chance variation. It could also be due to a poor assignment of treatments. For example, the right side/left side assignment is vulnerable to bias due to soil fertility, or drainage that is different from one side of the field to the other. The checkerboard assignment is also susceptible to fertility, drainage or other gradients.

- **Random Assignment**

  What if chance is used to assign varieties to plots? How, physically, would you randomly assign varieties to plots? Come up with a randomization scheme to assign variety A to 18 plots and variety B to the remaining 18 plots. Record your assignments in the table below.

Randomized Assignment

	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

Once you have completed your random assignment, ask your instructor for “The Truth” - this sheet gives the yield for each plot using either variety. “The Truth” was used to fill in the yields for the plots in the convenience and alternating patterns you looked at earlier. In general, "The Truth" is not available since it requires knowing what would happen to the same plot of land using each of the treatments.

Write down the yields for your random assignment - if you have an A in the row 1, column 1 plot then you would put down 130 whereas if you have a B in the row 1, column 1 plot you would put down 118 for the yield. Repeat for all squares. This gives you 18 A yields and 18 B yields. Based on this assignment, at random, did you find a significant difference in mean yield between the two corn varieties?

Share and discuss your results. Examine "The Truth" more closely. Which variety appears to have the larger yield? By how much?

THE TRUTH

A = 130 B = 118	A = 149 B = 137	A = 139 B = 127	A = 167 B = 155	A = 149 B = 137	A = 157 B = 145
A = 149 B = 137	A = 133 B = 121	A = 152 B = 140	A = 143 B = 131	A = 159 B = 147	A = 148 B = 136
A = 141 B = 129	A = 156 B = 144	A = 137 B = 125	A = 158 B = 146	A = 144 B = 132	A = 160 B = 148
A = 150 B = 138	A = 142 B = 130	A = 155 B = 143	A = 148 B = 136	A = 164 B = 152	A = 145 B = 133
A = 139 B = 127	A = 155 B = 143	A = 139 B = 127	A = 159 B = 147	A = 149 B = 137	A = 165 B = 153
A = 155 B = 143	A = 138 B = 126	A = 150 B = 138	A = 149 B = 137	A = 157 B = 145	A = 148 B = 136

Suggested Solution:

- **Convenience Assignment**

  Using a two independent sample analysis to compare the mean yields of the two varieties the value of the t-test statistic is 1.17 with an associated two sided P-value of 0.25. The P-value is the same whether you use the pooled or non-pooled option on the TI-83. If you are using the conservative degrees of freedom, $\min(n_1 - 1, n_2 - 1) = 17$, the P-value would be 0.26. Although variety A has a slightly larger mean yield, there is not a statistically significant difference between the sample mean yields for the two varieties.

- **Alternating Assignment**

  Using a two independent sample analysis to compare the mean yields of the two varieties the value of the t-test statistic is -1.20 with an associated two sided P-value of 0.24. The P-value is the same whether you use the pooled or non-pooled option on the TI-83. If you are using the conservative degrees of freedom, $\min(n_1 - 1, n_2 - 1) = 17$, the P-value would be 0.25. Although variety B has a slightly larger mean yield, there is not a statistically significant difference between the sample mean yields for the two varieties.

- **Random Assignment**

  How one randomly assigns varieties to plots is a good class discussion question. Some students might suggest flipping a coin for each plot; heads = A and tails = B. This is random but will not assure 18 plots with variety A and 18 with variety B.

  One way to randomly assign the varieties to the plots is to use a die.

  - Roll the die, this will give the row number for the plot
  - Roll the die again, this will give the column number for the plot
  - Assign variety A to the plot with the (row,column) numbers from above
  - Repeat the steps above until 18 plots have variety A
  - Fill in the remaining 18 plots with variety B

  Another way to randomly assign the varieties to plots is to use the TI-83 to generate a random assignment. Essentially what we want to do is to select 18 of the 36 plots at random to receive variety A. The remaining 18 plots (they are chosen at random by default) will receive variety B. To do this first label the plots sequentially from 1 to 36. Then using the TI-83 calculator;

– Put the numbers 1, 2, 3, . . . , 36 in L1.

– Generate 36 uniform random digits in L2.
Math → PRB → 1:rand → ENTER
rand**(36) → STO→ L2**

– Arrange L2 in ascending order while carrying the entries from L1 along.
2nd → LIST → OPS → 1:SortA(→ ENTER
SortA(**L2, L1) → ENTER**

– Read off the first 18 numbers in list L1. These plot numbers will receive variety A. The remaining plot numbers will receive variety B.

Example Randomization with yields

A 130	A 149	A 139	A 167	B 137	A 157
A 149	B 121	B 140	B 131	A 159	B 136
A 141	A 156	A 137	A 158	A 144	A 160
A 150	B 130	B 143	B 136	B 152	B 133
B 127	A 155	B 127	A 159	A 149	B 153
B 143	B 126	B 138	A 149	B 145	B 136

Summary	n	mean	std. dev
A	18	150.4	9.49
B	18	136.3	8.74

Using a two independent sample analysis to compare the mean yields of the two varieties the value of the t-test statistic is 4.64 with an associated two sided P-value that is virtually zero. The P-value is the same whether you use the pooled or non-pooled option on the TI-83. Even using the conservative degrees of freedom, $\min(n_1 - 1, n_2 - 1) = 17$, the P-value is virtually zero. Varieties have different mean yields and that difference is statistically significant.

Closer examination of "THE TRUTH" reveals that variety A has a yield that is 12 higher than variety B on every plot. The true difference in yield between variety A and variety B is 12 bushels per acre.

3 Discussion

Assignment by convenience or using an alternating pattern failed to uncover the true difference between the two varieties. "THE TRUTH" was set up in such a way that the convenience pattern and alternating pattern would mislead the experimenter. If you look closely at "THE TRUTH" you will see that there are alternating high/low yield gradients running diagonally across the field. By planting one variety on one side of the field, or in the alternating pattern, the superiority of variety A is hidden by these diagonal yield gradients. In real fields the truth is not known but non-random assignment of varieties to plots can mislead the experimenter in much the same way. The hidden patterns in real fields can confound the effects of the varieties.

Randomization, the random assignment of varieties to plots tends to take hidden patterns (or lurking variables) and spread their effects evenly across the treatment groups. This allows us to see the underlying truth most of the time. This disclaimer, "most of the time," is important. Even with randomization, we are not guaranteed to find a statistically significant difference even when a real difference does exist. In fact, the chance that a test of hypothesis can detect a difference when one exists is called the power of the test. By looking at the results of tests based on many random assignments, this activity can be used to simulate the power of the two sample t-test to detect a difference in mean yield of 12 bushels per acre. When this randomization activity was done by 40 AP statistics teachers at a short course, all but one of the teachers obtained a t-test statistic that was statistically significant. That is, the simulated power was 39 out of 40 or 97.5%.

4 More on Power

Let's look at power in a little more detail. What we would like to know is of all the possible randomizations of varieties to plots how many would produce a significant difference in sample mean yields? There are over 9 billion possible randomizations so enumerating all of them is out of the question. We can tackle this problem theoretically with some simplifying assumptions. For the two sample problem, it is easiest to look at power assuming normally distributed values with a common, and known variance. For the corn yield example we might assume that the yields for variety A are normally distributed with a mean μ_A and variance $\sigma^2 = 87$. Additionally, let's assume that the yields for variety B are normally distributed with a mean μ_B and variance $\sigma^2 = 87$. The value 87 for the population variance is obtained from the values reported in "THE TRUTH." We need to first establish what is a statistically significant difference. To do this we can use the 68-95-99.7 (or empirical) rule. Recall that approximately 95% of normally distributed values are within 2 standard deviations of the mean. So any difference whose absolute value is greater than 2 standard deviations is statistically significant at approximately the 5% level. We have the variances for individual yields but we need the variance (to get the standard deviation) of the difference in sample mean yields.

Sample mean yields (n=18) for variety A will be normally distributed with a center at

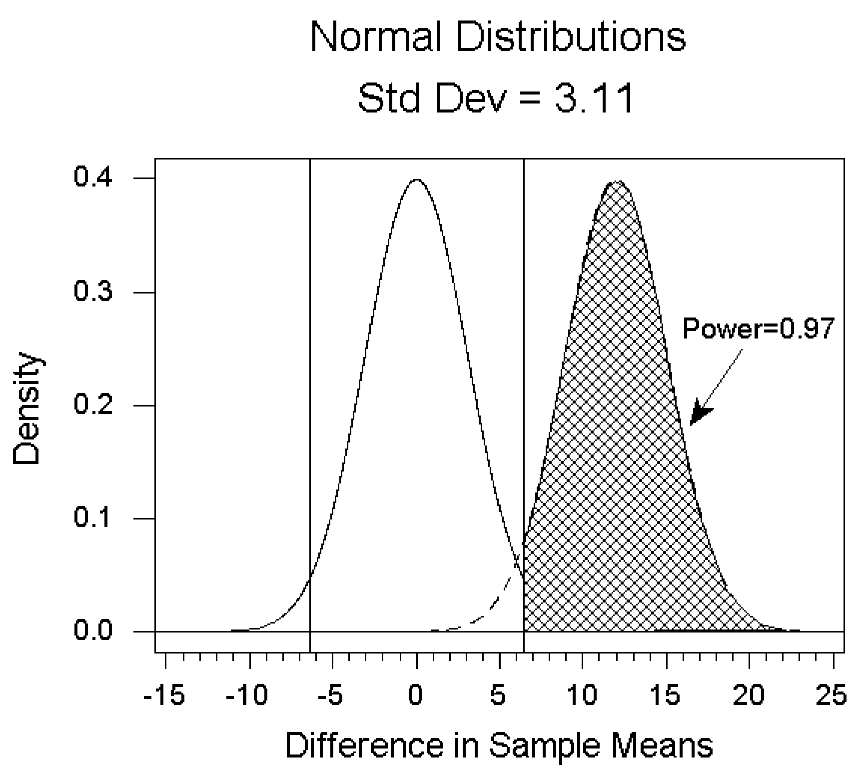
μ_A and a variance $\frac{\sigma^2}{n} = \frac{87}{18} = 4.833$. Similarly, sample mean yields (n=18) for variety B will be normally distributed with a center at μ_B and variance $\frac{\sigma^2}{n} = \frac{87}{18} = 4.833$. The difference in sample mean yields will be normally distributed with a center at $\mu_A - \mu_B$ and a variance of $\frac{\sigma^2}{n} + \frac{\sigma^2}{n} = \frac{87}{18} + \frac{87}{18} = 9.667$. Thus the standard deviation for the difference in two sample mean yields (n=18) is $\sqrt{9.667} = 3.11$. Any absolute difference in sample mean yields larger than two standard deviations (6.22) would be considered statistically significant.

To calculate the power all we would need to do is to compute the probability of getting a difference in sample mean yields that is less than -6.22 or greater than 6.22 **when we assume the true difference in means** $\mu_A - \mu_B = 12$. This is just the probability that a normal random variable with mean 12 and standard deviation 3.11 takes on a value less than -6.22 or greater than 6.22. We can obtain the standardized values

$$z_1 = \frac{-6.22 - 12}{3.11} = -5.86$$

$$z_2 = \frac{6.22 - 12}{3.11} = -1.86$$

The normal cumulative distribution function (cdf) for z_1 is zero and so contributes nothing to the power calculation. The cdf for z_2 is 0.03, so the chance of being greater than $z_2 = -1.86$, and thus the power, is 1 - .03 = .97. The computation of the power is illustrated in the figure below.



Power is actually a function of how big a difference you want to detect. In the calculation above, the true difference of 12 will be picked up most of the time by a two independent sample test when randomization is used to assign varieties to plots. The power will be much lower for smaller true differences. You can adjust "THE TRUTH" so that variety A beats variety B by say 6 bushels. You will find that the power as calculated above (think about moving the right hand normal curve in the figure above so that it is centered at 6 instead of 12) is less than before (around 0.50). Power is clearly a function of the size of the true difference. Procedures have more power to detect large differences than small differences. Power is also affected by sample size. We know that larger sample sizes are good because they reduce the variation in the sample mean. It is nice to know that larger sample sizes also provide more power for much the same reason. Think about how the graph above would change if the sample size, the number of plots receiving each variety increased.