Designs for experiments

by Leshun Xu

27 Apr 2019

Historically, the agricultural experiment is the earliest application area on Design of Experiments. This document is an example about how designs can affect experiments. It includes a method for weighing products with better precision, a test on whether the nitrogen fertilizer affects the crops, and an A/B test for two strategies. Currently, there are many packages that can be found in RCAN. However, in this example, only "pwr" package is used for A/B testing. For more details on other packages, please refer to CRAN Task View: Design of Experiments (DoE) & Analysis of Experimental Data and Design of Experiments in R (the invited talk at The R User Conference 2011) by Ulrike Groemping.

Weighing three balls with a new balance

The "balls" here can be livestocks, crops, and any other products in industries. The goal is to weight them with smaller variation, which means better precision. This idea was inspired by an example on Wikipedia.

Let's suppose the variance of measures is σ^2 (of cause, in i.i.d. situations). If we use the balance in Figure 1, and we weigh one object (i.e. one ball) at a time, then the variance is certainly σ^2.

If we use the design introduced in Wikipedia (i.e. let $B_i, (i = 1, 2, 3)$ stand for three balls, and $Y_j, j = 1, 2, 3, 4$, are measued differences described in the following table), then the weight of each ball can be estimated by the following formulas:

Steps	Left Pan	Right Pan	Measured Difference
step 1	B_1, B_2, B_3	empty	$Y_1 = (B_1 + B_2 + B_3) - 0$
step 2	B_1, B_3	B_2	$Y_2 = (B_1 + B_3) - B_2$
step 3	B_2, B_3	B_1	$Y_2 = (B_2 + B_3) - B_1$
step 4	B_1, B_2	B_3	$Y_2 = (B_1 + B_2) - B_3$

$$\begin{aligned}
\text{estimated weight of } B_1 &= \frac{Y_1 + Y_2 - Y_3 + Y_4}{4}, \\
\text{estimated weight of } B_2 &= \frac{Y_1 - Y_2 + Y_3 + Y_4}{4}, \\
\text{estimated weight of } B_3 &= \frac{Y_1 + Y_2 + Y_3 - Y_4}{4}.
\end{aligned}$$

Now, since measures are i.i.d., the variance of the estimation becomes $\sigma^2/4$, which implies higher precision.

However, this design is hard to use for practice problems when the number of objects is greater, when the objects are hard to be measured more than once, or when the balance limitation is surpassed. Therefore, I suggest the following newly-designed balance:

Let's suppose the "left arm" is k times the length of the "right arm" in this new balance. Then it should give us k^2 times as much precision than the old balance does.

Nitrogen fertilizer for corn.

This example is from the book *Experimental Design and Analysis* by Howard J. Seltman, and the corn data set is downloaded from his website. The goal is to evaluate whether the nitrogen fertilizer benefits the weight of corn.

Figure 1: Figure 1: The balance (Refer to Wikipedia)

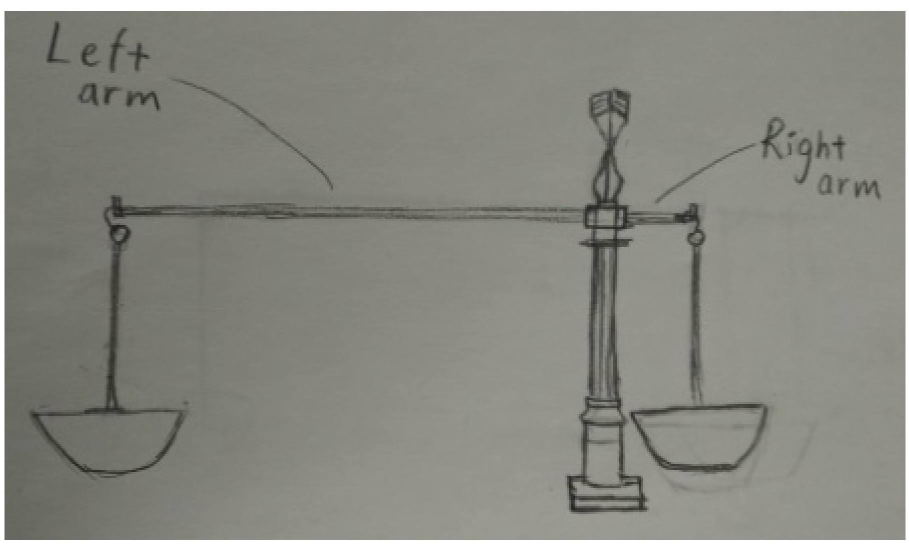

Figure 2: Figure 2: The newly-designed balance (Art work by Sophia)

We assume the effects of nitrogen fertilizer is linear to the corn weight, i.e. $\mathbb{E}(weight|nitrogen) = \beta_0 + \beta_1 nitrogen$. Then the statistical hypotheses are

H0 : $\beta_1 = 0$.

H1 : $\beta_1 \neq 0$.

By using the simple linear regression, we have the following summary for the fitted model.

```
##
## Call:
## lm(formula = weight ~ nitrogen, data = corns)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -63.19 -33.41 -11.38  31.38 112.69
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  84.8214    18.1158   4.682 0.000114 ***
## nitrogen      5.2686     0.2992  17.610 1.87e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 50.06 on 22 degrees of freedom
## Multiple R-squared:  0.9338, Adjusted R-squared:  0.9307
## F-statistic: 310.1 on 1 and 22 DF,  p-value: 1.869e-14
```

The above shows β_1 is estimated as 5.2686, and the nitrogen fertilizer significantly and positively contributes to the weight of corn. Therefore, the benefits of the nitrogen fertlizer cannot be ignored (i.e. H0 is rejected) by farmers.

R package "pwr" for A/B testing

We repeatedly assume that there are two strategies for planting corn, with Strategy A using nitrogen fertilizer, and Strategy B not using nitrogen fertilizer. Let's suppose 1000 farming plots were examined and measured, and 100 of them were found to be converted, gaining weight. Details are listed in the following table:

Strategies	Number of Plots	Number of Converted Plots
A	520	62
B	480	38

The statistical hypotheses are:

H0: there is **no** difference between Strategy A and Strategy B.

H1: there **exits** a difference between Strategy A and Strategy B.

Since the difference between two strategies could be either positive or negative, we will use the two-tail test in the "prop.test" function. We have the following output:

```
##
##  2-sample test for equality of proportions with continuity
##  correction
##
## data:  c(62, 38) out of c(520, 480)
## X-squared = 4.0175, df = 1, p-value = 0.04503
```

```
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.001193507 0.078934698
## sample estimates:
##     prop 1     prop 2
## 0.11923077 0.07916667
```

As we can see, the p-value is 0.045. It means under the 5% significant level, so we can reject the hypothesis that conversion rates are equal, and take Strategy A for having a better harvest.

The "pwr" package can also provide recommendations for the sample size. For example, we want to use 520 plots as the sample for Strategy A, and we set the Type I error to less than 5%, the Type II error to less than 15%, and assume the effect size to be h = 0.2, in the "pwr.2p2n.test" function. Then, we have the following output:

```
##
##      difference of proportion power calculation for binomial distribution (arcsine transformation)
##
##               h = 0.2
##              n1 = 520
##              n2 = 394.9344
##       sig.level = 0.05
##           power = 0.85
##     alternative = two.sided
##
## NOTE: different sample sizes
```

Based on the above, the suggested sample size for Strategy B is at least 395.