UNIT III

Statistical Experiments and Significance Testing

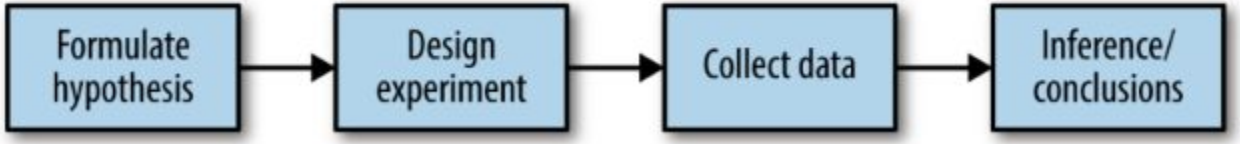


Figure 3-1. The classical statistical inference pipeline

A/B Test

An A/B test is an experiment with two groups to establish which of two treatments, products, procedures, or the like is superior.

Often one of the two treatments is the standard existing treatment.

If a standard treatment is used,it is called the control.

A typical hypothesis is that a new treatment is better than the control.

Examples of A/B Test

- Testing two soil treatments to determine which produces better seed germination
- Testing two therapies to determine which suppresses cancer more effectively
- Testing two prices to determine which yields more net profit
- Testing two web headlines to determine which produces more clicks.
- Testing two web ads to determine which generates more conversions

Hypothesis Tests

Hypothesis tests, also called significance tests, Their purpose is to help you learn whether random chance might be responsible for an observed effect.

Null hypothesis : The hypothesis that chance is to blame.

Alternative hypothesis: Counterpoint to the null (what you hope to prove).

One-way test: Hypothesis test that counts chance results only in one direction.

Two-way test: Hypothesis test that counts chance results in two directions.

Null Hypothesis

Logic : In a properly designed A/B test, you collect data on treatments A and B in such a way that any observed difference between A and B must be due to either:

- Random chance in assignment of subjects
- A true difference between A and B.

“Given the human tendency to react to unusual but random behavior and interpret it as something meaningful and real, in our experiments we will require proof that the difference between groups is more extreme than what chance might reasonably produce.”

Baseline assumption that the treatments are equivalent, and any difference between the groups is due to chance. This baseline assumption is termed the null hypothesis.

Our hope, is that we can in fact prove the null hypothesis wrong

Alternative Hypothesis

Null = “no difference between the means of group A and group B”; alternative = “A is different from B” (could be bigger or smaller)

- Null = “A ≤ B”; alternative = “A > B”
- Null = “B is not X% greater than A”; alternative = “B is X% greater than A”

One-Way Versus Two-Way Hypothesis Tests

Often in an A/B test, you are testing a new option (say, B) against an established default option (A), and the presumption is that you will stick with the default option unless the new option proves itself definitively better.

One-way Test

you want a hypothesis test to protect you from being fooled by chance in the direction favoring B.

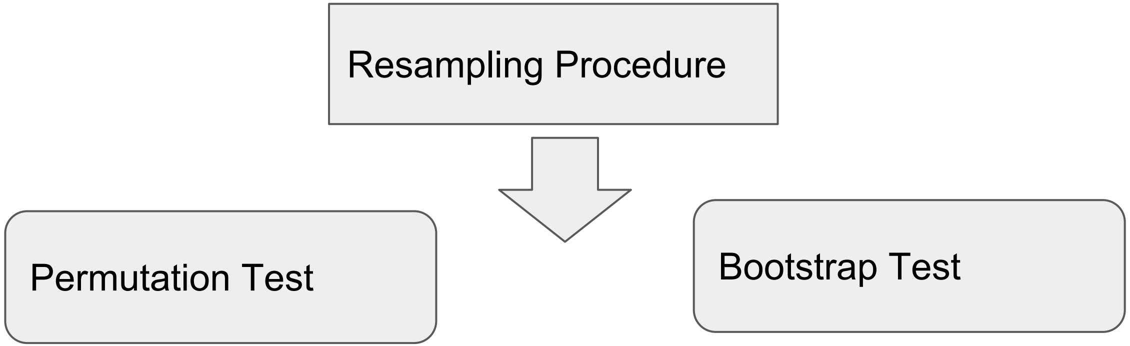
You don’t care about being fooled by chance in the other direction, because you would be sticking with A unless B proves definitively better.

Two way test

If you want a hypothesis test to protect you from being fooled by chance in either direction, the alternative hypothesis is bidirectional (A is different from B; could be bigger or smaller). In such a case, you use a two-way (or two-tail) hypothesis. This means that extreme chance results in either direction count toward the p-value.

Resampling

Resampling in statistics means to repeatedly sample values from observed data,with a general goal of assessing random variability in a statistic.



Permutation Test

The procedure of combining two or more samples together and randomly reallocating the observations to resamples.

The first step in a permutation test of a hypothesis is to combine the results from groups A and B (and, if used, C, D,...).

We then test that hypothesis by randomly drawing groups from this combined set and seeing how much they differ from one another.

Permutation Test : Steps

1. Combine the results from the different groups into a single data set.

2. Shuffle the combined data and then randomly draw (without replacement) a resample of the same size as group A (clearly it will contain some data from the other groups).

3. From the remaining data, randomly draw (without replacement) a resample of the same size as group B.

4. Do the same for groups C, D, and so on. You have now collected one set of resamples that mirror the sizes of the original samples.

5. Whatever statistic or estimate was calculated for the original samples (e.g., difference in group proportions), calculate it now for the resamples, and record; this constitutes one permutation iteration.

6. Repeat the previous steps R times to yield a permutation distribution of the test statistic.

Example: Web Stickiness

A company selling a relatively high-value service wants to test which of two web presentations does a better selling job.

Due to the high value of the service being sold, sales are infrequent and the sales cycle is lengthy; it would take too long to accumulate enough sales to know which presentation is superior.

So the company decides to measure the results with a proxy variable, using the detailed interior page that describes the service.

Proxy Variable

A proxy variable is one that stands in for the true variable of interest, which may be unavailable, too costly, or too time-consuming to measure.

One potential proxy variable for our company is the number of clicks on the detailed landing page. A better one is how long people spend on the page. It is reasonable to think that a web presentation (page) that holds people’s attention longer will lead to more sales. Hence, our metric is average session time, comparing page A to page B.

Exhaustive and Bootstrap Permutation Tests

Two variants of the permutation test:

- An exhaustive permutation test
- A bootstrap permutation test

In an exhaustive permutation test,we actually figure out all the possible ways data could be divided.

In a bootstrap permutation test, the draws outlined in steps 2 and 3 of the random permutation test are made with replacement instead of without replacement.

Statistical Significance and p-Values

P value is probability of null hypothesis being true.

Null Hypothesis: An assumption that treat everything equal and similar

Hypothesis testing : The process of accepting or rejecting null Hypothesis

Collect Data

Significance level(5 % OR 1 %)[The probability that a random value of a statistic will lie in the critical region is called level of significance]

Test to get P value (T test,chi Square,Anova,Z test)

Either u accept or reject null hypothesis

Error in testing of Hypothesis

Type I Error: A true hypothesis is rejected .i.e. When the difference between the sample value and hypothetical value exceeds the confidence limit. The error can be minimized by increasing the confidence interval.

Type II Error: when a false hypothesis is accepted.i.e. When the difference between the sample value and hypothetical value lie within the confidence limit.The error can be minimized by decreasing the confidence interval.

Permutation Test

Concluding Remark :

- In a permutation test, multiple samples are combined and then shuffled.
- The shuffled values are then divided into resamples, and the statistic of interest is calculated.
- This process is then repeated, and the resampled statistic is tabulated.
- Comparing the observed value of the statistic to the resampled distribution allows you to judge whether an observed difference between samples might occur by chance.

Why Annova

To this point we have been comparing two populations

Of course limiting ourselves to the comparison of two populations is limiting

What if we wish to compare the means of more than two populations ?

What if we wish to compare populations each containing several levels or subgroup

ANOVA: ANalysis Of VAriance

Suppose we want to compare three sample means to see if a difference exists somewhere among them

So all three means come from a common population ?

Is one mean so far away from the other two that it is likely not from the same population?

Or

All three are far apart that they all likely come from unique population ?

Anova Steps

Steps

1. State null & alternative hypotheses

2. State Alpha
3. Calculate degrees of Freedom
4. State decision rule

5. Calculate test statistic
- Calculate variance between samples
- Calculate variance within the samples
- Calculate F statistic

1. State null & alternative hypotheses

$$H_0 = \mu_1 = \mu_2 ... = \mu_i$$

H0 : all sample means are equal

$$H_a = \text{not all of the } \mu_i \text{ are equal}$$

At least one sample has different mean

Steps/ Variance between Samples

2. State Alpha i.e 0.05

3. Calculate degrees of Freedom

K-1 & n-1

k= No of Samples,

n= Total No of observations

4. State decision rule

If calculated value of F >table value of F, reject Ho

5. Calculate test statistic

1. Calculate the **mean** of each sample.
2. Calculate the **Grand average**
3. Take the difference between means of various samples & grand average.
4. Square these deviations & obtain total which will give sum of squares between samples **(SSC)**
5. Divide the total obtained in step 4 by the degrees of freedom to calculate the mean sum of square between samples **(MSC)**.

Calculating Variance within Samples

1. Calculate **mean** value of each sample
2. Take the deviations of the various items in a sample from the mean values of the respective samples.
3. Square these deviations & obtain total which gives the sum of square within the samples **(SSE)**
4. Divide the total obtained in 3rd step by the degrees of freedom to calculate the mean sum of squares within samples **(MSE).**

The mean sum of squares

Calculation of **MSC**- **M**ean sum of **S**quares between samples

$$MSC = \frac{SSC}{k-1}$$

Calculation of **MSE** Mean Sum Of Squares within samples

$$MSE = \frac{SSE}{n-k}$$

k= No of Samples, n= Total No of observations

Calculation of F statistic

$$F = \frac{Variability\ between\ groups}{Variability\ within\ groups}$$

$$\text{F- statistic} = \frac{MSC}{MSE}$$

Compare the F-statistic value with F(critical) value which is obtained by looking for it in F distribution tables against degrees of freedom. The calculated value of F > table value H0 is rejected

Example- one way ANOVA

Example: 3 samples obtained from normal populations with equal variances. Test the hypothesis that sample means are equal

8	7	12
10	5	9
7	10	13
14	9	12
11	9	14

1.Null hypothesis –
No significant difference in the means of 3 samples

2. State Alpha i.e 0.05

3. Calculate degrees of Freedom
k-1 & n-k = 2 & 12

4. State decision rule
Table value of F at 5% level of significance for d.f 2 & 12 is 3.88
The calculated value of F > 3.88 ,H0 will be rejected

5. Calculate test statistic

X1	X2	X3
8	7	12
10	5	9
7	10	13
14	9	12
11	9	14
Total 50 M1= 10	40 M2 = 8	60 M3 = 12

$$\text{Grand average} = \frac{10+8+12}{3} = 10$$

Variance BETWEEN samples (M1=10, M2=8,M3=12)

Sum of squares between samples (SSC) =

n1 (**M1** – Grand avg)^2 + n2 (**M2**– Grand avg)^2 + n3(**M3**– Grand avg)^2

5 (10 - 10)^2 + 5 (8 - 10)^2 + 5 (12 - 10)^2 = 40

Calculation of **M**ean sum of **S**quares between samples (**MSC**)

$$MSC = \frac{SSC}{k-1} = \frac{40}{2} = 20$$

k= No of Samples, n= Total No of observations

Variance WITH IN samples (M1=10, M2=8,M3=12)

X1	$(X1 - M1)^2$	X2	$(X2 - M2)^2$	X3	$(X3 - M3)^2$
8	4	7	1	12	0
10	0	5	9	9	9
7	9	10	4	13	1
14	16	9	1	12	0
11	1	9	1	14	4
	30		16		14

Sum of squares within samples (SSE) = 30 + 16 +14 = 60

Calculation of Mean Sum Of Squares within samples (**MSE)**

$$MSE = \frac{SSE}{n-k} = \frac{60}{12} = 5$$

Calculation of ratio F

$$F = \frac{Variability\ between\ groups}{Variability\ within\ groups}$$

F- statistic = $\frac{MSC}{MSE}$ = 20/5 =4

The Table value of F at 5% level of significance for d.f 2 & 12 is 3.88
The calculated value of F > table value
H0 is rejected. Hence there is significant difference in sample means

Short cut method -

X1	$(X1)^2$	X2	$(X2)^2$	X3	$(X3)^2$
8	64	7	49	12	144
10	100	5	25	9	81
7	49	10	100	13	169
14	196	9	81	12	144
11	121	9	81	14	196
Total 50	530	40	336	60	734

Total sum of all observations = 50 + 40 + 60 = 150

Correction factor = $T^2 / N = (150)^2 / 15 = 22500/15 = 1500$

Total sum of squares= 530+ 336+ 734 – 1500= 100

Sum of square b/w samples=$(50)^2/5 + (40)^2/5 + (60)^2/5 - 1500 = 40$

Sum of squares within samples= 100-40= 60