Applied Data Science

Session 11-12: Hypothesis Testing, A/B Testing

Dr. Soharab Hossain Shaikh

1

Prerequisite

- In statistics **"population"** refers to the total set of observations that can be made.
- e.g., if we want to calculate average height of humans present on the earth, **"population"** will be the "total number of people actually present on the earth".
- A **sample,** on the other hand, is a set of data collected/selected from a pre-defined procedure.
- For our example above, it will be a small group of people selected randomly from some parts of the earth.

2

Prerequisite

- To draw inferences from a sample by validating a hypothesis it is necessary that the sample is **random**.
- For instance, in our example above if we select people randomly from all regions(Asia, America, Europe, Africa etc.)on earth, our estimate will be close to the actual estimate and can be assumed as a sample mean, whereas if we make selection let's say only from the United States, then our average height estimate will not be accurate but would only represent the data of a particular region (United States).
- Such a sample is then called a biased sample and is not a representative of "population".

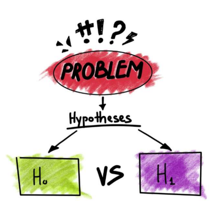
3

Prerequisite

- Another important aspect to understand in statistics is **"distribution"**.
- When "population" is infinitely large it is improbable to validate any hypothesis by calculating the mean value or test parameters on the entire population.
- In such cases, a population is assumed to be of some type of a distribution.

4

Hypothesis Testing



Hypothesis testing is a critical tool in inferential statistics, for determining what the value of a ***population parameter*** could be.

We often draw this conclusion based on a ***sample data analysis***.

5

Hypothesis Testing

- The basis of hypothesis testing has two attributes:

(a) **Null Hypothesis** (H_0)and

(b) **Alternative Hypothesis** (H_1or H_A)

- The null hypothesis is, in general, the **boring stuff** i.e. it assumes that nothing interesting happens/happened.
- The alternative hypothesis is, **where the action is** i.e. some observation/ phenomenon is real (i.e. not a fluke) and statistical analysis will give us more insights on that.

6

What is the Process?

- Statisticians take a pessimistic sort of view and start with the Null hypothesis, and compute some sort of test-statistic in the sample data.
- It is given by, $\frac{Best\ Estimate - Hypothesized\ Estimate}{Standard\ Error\ of\ Estimate}$
- Here, the '*best estimate*' comes from the sample e.g. sample mean or proportion of some data in the sample.
- Standard error represents the variability in this estimate and often depends on the variance and sample size.

7

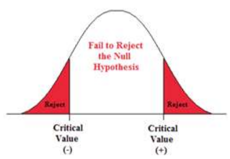
Hypothesis Testing and p-value

- In Null hypothesis significance testing, the p-value is the probability of obtaining test results at least as extreme as the result actually observed, under the assumption that the null hypothesis is correct.
- This chance — probability value of observing the test-statistic — is the so-called **p-value.**
- The smaller the **p-value**, the stronger the evidence against the Null hypothesis. If p-value is less than the level of significance (α), we should **reject the Null hypothesis.**
- **p-values** are expressed as decimals although it may be easier to understand what they are if you convert them to a percentage. For example, **a p value** of 0.0254 is 2.54%.

8

Critical Value

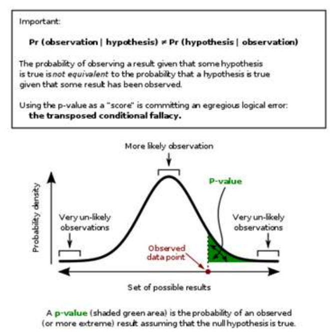
The p-value is the probability of observing the test-statistic, as is, given the Null hypothesis is true.
This probability is calculated under the assumption of a certain probability distribution (that the test statistic is generated from).



If this particular value is very small (less than a pre-determined **Critical Value**), we can reject the Null hypothesis in favour of the alternative.

9

p-value



10

Critical Value

- A critical value is a point (or points) on the scale of the test statistic beyond which we reject the null hypothesis, and, is derived from the level of significance α of the test.
- **Critical value can tell us, what is the probability of two sample means belonging to the same distribution.**
- **Higher, the critical value means lower the probability of two samples belonging to same distribution**.
- The general critical value for a two-tailed test is **1.96,** which is based on the fact that **95%** of the area of a normal distribution is within 1.96 standard deviations of the mean.

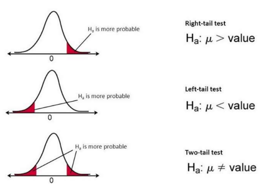
11

Critical Value

- Critical values can be used to do hypothesis testing in following way
- 1. Calculate test statistic
- 2. Calculate critical values based on significance level - alpha (α)
- 3. Compare test statistic with critical values.
- 4. If the test statistic is falling in the critical region, we reject the Null Hypothesis in favour of the Alternate Hypothesis.

12

One-tailed / Two-tailed Tests

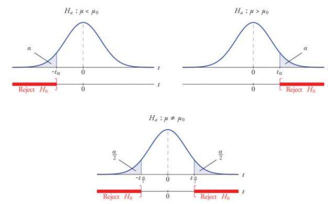


In some situations, the hypothesis deals with questions in the form of "***x* is greater than *y***" or "***x* is lesser than *y***".
In those cases, only one side of the probability distributions have to be checked for and we call them 'one-sided' or 'one-tailed' test.

In other situation, we have to use both sides of the probability distribution.
That is called 'two-sided' or 'two-tailed' test.
In these situations, the alternative hypothesis is generally expressed in the form "***x* is not equal to *y***".

13

Test Outcome



- Hypothesis tests are based on the notion of critical regions.
- The Null Hypothesis is **rejected** in favour of the alternate if the test statistic falls in the critical region.
- Otherwise, we reject the alternate hypothesis in favour of the Null hypothesis.

14

Steps – Hypothesis Testing

- Step 1 State the hypotheses and identify the claim.
- Step 2 Find the critical value(s) from the appropriate table.
- Step 3 Compute the test value.
- Step 4 Make the decision to reject or not reject the null hypothesis.
- Step 5 Summarize the results.

15

Z-test

In a z-test, the sample is assumed to be normally distributed.

A z-score is calculated with population parameters such as **"population mean"** and **"population standard deviation"** and **is used to validate a hypothesis that the sample drawn belongs to the same population.**

Null: Sample mean is same as the population mean

Alternate: Sample mean is not same as the population mean

The statistics used for this hypothesis testing is called z-statistic, the score for which is calculated as follows:

Sample mean | Population mean

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Population standard deviation | Number of samples

If the test statistic is lower than the critical value, we fail to reject the Null hypothesis or else reject it.

16

T-test

In statistics, the term "t-test" refers to the hypothesis test in which the test statistic follows a Student's t-distribution.

A t-test is used to compare the mean of two given samples.

Like a z-test, a t-test also assumes a normal distribution of the sample.

A t-test is used when the population parameters (mean and standard deviation) are not known.

There are three versions of t-test

1. Independent samples t-test which compares mean for two groups
2. Paired sample t-test which compares means from the same group at different times
3. One sample t-test which tests the mean of a single group against a known mean.

17

T-test

T-test is used to check whether two data sets are significantly different from each other or not.

One of the variants of the t-test is the one-sample t-test which is used to determine if the sample is significantly different from the population.

The formula for a one-sample t-test is expressed using the observed sample mean, the theoretical population means, sample standard deviation, and sample size. Mathematically, it is represented as,

t = (x̄ – μ) / (s / √n)

where

- x̄ = Observed Mean of the Sample
- μ = Theoretical Mean of the Population
- s = Standard Deviation of the Sample
- n = Sample Size

18

T-test

- In case statistics of two samples are to be compared, then a two-sample t-test is to be used and its formula is expressed using respective sample means, sample standard deviations, and sample sizes.
- Mathematically, it is represented as,

$t = (\bar{x}_1 - \bar{x}_2) / \sqrt{[(s^2_1 / n_1) + (s^2_2 / n_2)]}$

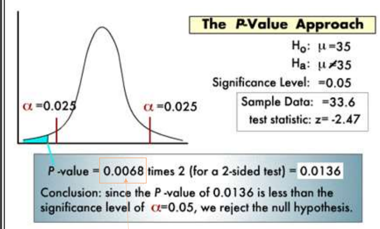
Where,

- $\bar{x}_1$ = Observed Mean of 1st Sample
- $\bar{x}_2$ = Observed Mean of 2nd Sample
- s_1 = Standard Deviation of 1st Sample
- s_2 = Standard Deviation of 2nd Sample
- n_1 = Size of 1st Sample
- n_2 = Size of 2nd Sample

If the test statistic is lower than the critical value, we fail to reject the Null hypothesis or else reject it.

19

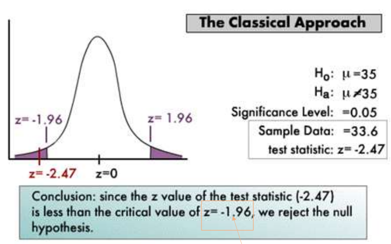
p-value Approach



From the Standard Normal table for z=-2.47

20

Classical Approach



From the Standard Normal table

21

Standard Normal Table

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

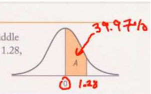
Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
-1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
-1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
-1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
-1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
-1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
-1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811

22

Standard Normal Table

Normal Curve (z) Table

The normal curve table gives only the percentage of data starting from the middle ($z = 0$), out to whatever z score you look up. For instance, if you look up $z = 1.28$, you get .3997. This means .3997 or 39.97% of the data in the normal curve is found between $z = 0$ and $z = 1.28$.



					Normal					
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706

23

How to Find Probabilities for Z with the Z-Table?

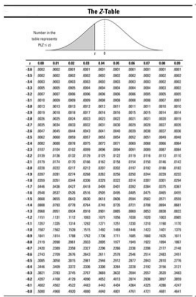
- You can use the **Z-table** to find a full set of "less-than" probabilities for a wide range of *z*-values. To use the *Z*-table to find probabilities for a **statistical sample** with a standard normal (*Z*-) distribution, do the following:

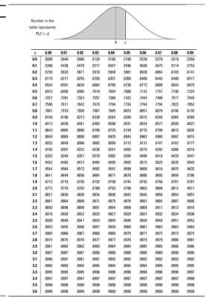
1. Go to the row that represents the ones digit and the first digit after the decimal point (the tenths digit) of your *z*-value.
2. Go to the column that represents the second digit after the decimal point (the hundredths digit) of your *z*-value.
3. Intersect the row and column from Steps 1 and 2. This result represents $p(Z < z)$, the probability that the random variable *Z* is less than the value *z* (also known as the percentage of *z*-values that are less than the given *z* value).

24

Computing p-Value

- For example, suppose you want to find $p(Z < 2.13)$.
- Using the Z-table below, find the row for 2.1 and the column for 0.03.
- Intersect that row and column to find the probability: 0.9834.
- Therefore $p(Z < 2.13) = 0.9834$.





25

Computing p-Value

- Noting that the total area under any normal curve (including the standardized normal curve) is 1, it follows that

$p(Z < 2.13) + p(Z > 2.13) = 1$.

Therefore, $p(Z > 2.13) = 1 - p(Z < 2.13)$ which equals 1 – 0.9834 which equals 0.0166.

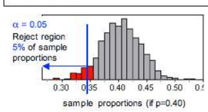
- Suppose you want to look for $p(Z < -2.13)$. You find the row for –2.1 and the column for 0.03. Intersect the row and column and you find 0.0166; that means $p(Z < -2.13)=0.0166$. Observe that this happens to equal $p(Z>+2.13)$.
- The reason for this is because the normal distribution is symmetric.
- So the tail of the curve below –2.13 representing $p(Z < -2.13)$ looks exactly like the tail above 2.13 representing $p(Z > +2.13)$.

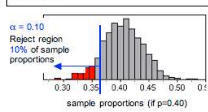
26

Significance Levels

Significance Level 5% versus 10%

α = 5%	α = 10%
The P-value is 7.8%. The P-value is **greater than** α.	The P-value is 7.8%. The P-value is **less than** α.
Conclusion: Data is NOT statistically significant. (Not rare in random sampling if Ho is true.)	**Conclusion:** Data is statistically significant. (Rare in random sampling if Ho is true.)
Fail to reject Ho.	**Reject Ho.**





27

Confidence Interval

A (1 - α) confidence interval for the population mean…

$$\bar{x} - |z_{\alpha/2}|\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + |z_{\alpha/2}|\frac{\sigma}{\sqrt{n}}$$

Lower limit Upper limit

When the population standard deviation (σ) is not known,

- we replace it by the ***sample standard deviation (s)***
- The z-statistic ($z_{\alpha/2}$) gets replaced by the ***t*-statistic ($t_{\alpha/2}$)**

A (1 - α) confidence interval for the population mean…

$$\bar{x} - |t_{\alpha/2}|\frac{s}{\sqrt{n}} < \mu < \bar{x} + |t_{\alpha/2}|\frac{s}{\sqrt{n}}$$

28

Hypothesis Testing in Model Validation

$$Sales = \beta_0 + \beta_1 Price + \beta_2 AdExp + \beta_3 PromExp$$

$$Sales = -25096.83 - 5055.27Price + 648.61AdExp + 1802.61PromExp$$

For every 1000 dollars increase in advertisement spending, the sales increase by 648.6 (~649) units, all other variables remaining at the same level.

Belief held by Salespeople…

For every 1000 dollars increase in advertising expenditure, the unit sales increase by 500 units.

29

Hypothesis Testing in Model Validation

Hypothesis testing needed to test this belief.

True beta values, ***fixed*** but ***unknown***

$$Sales = \beta_0 + \beta_1 Price + \beta_2 AdExp + \beta_3 PromExp$$

$$Sales = -25096.83 - 5055.27Price + 648.61AdExp + 1802.61PromExp$$

Estimates based on ***Sample data***

Hypothesis test needed to test whether β_2 is equal to 500

30

Hypothesis Testing in Model Validation

$$Sales = \beta_0 + \beta_1 Price + \beta_2 AdExp + \beta_3 PromExp$$

Could the true value of β_2 be 500?

H_0: $\beta_2 = 500$
H_A: $\beta_2 \neq 500$

Conclusion:

- Do not reject the Null hypothesis
- True value of β_2 could be 500
- We cannot reject the belief held by salespeople

31

Confidence Interval Approach to Hypothesis Testing

$$Sales = \beta_0 + \beta_1 Price + \beta_2 AdExp + \beta_3 PromExp$$

ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	197798832.8	65932944	40.56262	1.0848E-08
Residual	20	32509212.11	1625461		
Total	23	230308045			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-25096.83	24859.61131	-1.009542	0.324773	-76953.0734	26759.408
Price ($)	-5055.27	526.3995537	-9.603484	6.22E-09	-6153.32009	-3957.22
Adexp ('000$)	648.61214	209.0048787	3.103335	0.005602	212.635603	1084.5887
Promexp ('000$)	1802.611	392.8485427	4.588565	0.000178	983.143256	2622.0787

32

Confidence Interval Approach to Hypothesis Testing

Step 1 **: Formulate Hypothesis**

H_0: $\beta_2 = 500$
H_A: $\beta_2 \neq 500$

Step 2 **: Consider the 95% confidence interval for β_2**

= [212.6, 1084.6]

center point = 648.6

Conclusion:

- **Since 500 falls in the confidence interval, hence do not reject the Null hypothesis.**
- **We cannot reject the H_0 for any value that is in the confidence interval.**

33

p-Value and Model Coefficients

p-values and their importance in interpreting regression results

$$Sales = \beta_0 + \beta_1 Price + \beta_2 AdExp + \beta_3 PromExp$$

ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	197798832.8	65932944	40.56262	1.0848E-08
Residual	20	32509212.11	1625461		
Total	23	230308045			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-25096.83	24859.61131	-1.009542	0.324773	-76953.0734	26759.408
Price ($)	-5055.27	526.3995537	-9.603484	6.22E-09	-6153.32009	-3957.22
Adexp ('000$)	648.61214	209.0048787	3.103335	0.005602	212.635603	1084.5887
Promexp ('000$)	1802.611	392.8485427	4.588565	0.000178	983.143256	2622.0787

H_0: $\beta_0 = 0$
H_A: $\beta_0 \neq 0$

H_0: $\beta_1 = 0$
H_A: $\beta_1 \neq 0$

H_0: $\beta_2 = 0$
H_A: $\beta_2 \neq 0$

H_0: $\beta_3 = 0$
H_A: $\beta_3 \neq 0$

- Reject the Null hypothesis that $\beta_2 = 0$.
- Advertising expenditure is an important variable in explaining Sales.

34

Model Coefficients and p-Value

```
model = sm.OLS(y, X)
results = model.fit()
print(results.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       1.000
Model:                            OLS   Adj. R-squared:                  1.000
Method:                 Least Squares   F-statistic:                 4.020e+06
Date:                Thu, 06 Aug 2020   Prob (F-statistic):          2.83e-239
Time:                        13:04:02   Log-Likelihood:                -146.51
No. Observations:                 100   AIC:                             299.0
Df Residuals:                      97   BIC:                             306.8
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.3423      0.313      4.292      0.000       0.722       1.963
x1            -0.0402      0.145     -0.278      0.781      -0.327       0.247
x2            10.0103      0.014    715.745      0.000       9.982      10.038
==============================================================================
Omnibus:                        2.042   Durbin-Watson:                   2.274
Prob(Omnibus):                  0.360   Jarque-Bera (JB):                1.875
Skew:                           0.234   Prob(JB):                        0.392
Kurtosis:                       2.519   Cond. No.                         144.
==============================================================================
```

35

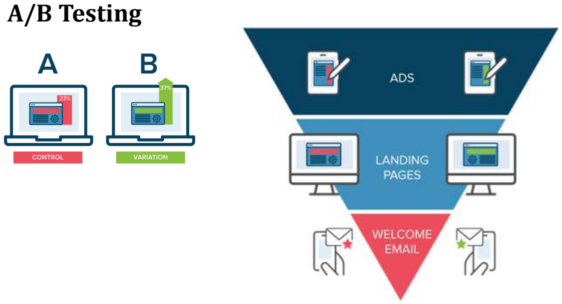
Model Coefficients and p-Value

```
ols_model = sm.OLS(y, X)
ols_results = ols_model.fit()
print(ols_results.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 TOTEMP   R-squared:                       0.995
Model:                            OLS   Adj. R-squared:                  0.992
Method:                 Least Squares   F-statistic:                     330.3
Date:                Thu, 06 Aug 2020   Prob (F-statistic):           4.98e-10
Time:                        13:04:03   Log-Likelihood:                -109.62
No. Observations:                  16   AIC:                             233.2
Df Residuals:                       9   BIC:                             238.6
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const      -3.482e+06    8.9e+05     -3.911      0.004    -5.5e+06   -1.47e+06
GNPDEFL       15.0619     84.915      0.177      0.863    -177.029     207.153
GNP           -0.0358      0.033     -1.070      0.313      -0.112       0.046
UNEMP         -2.0202      0.488     -4.136      0.003      -3.125      -0.915
ARMED         -1.0332      0.214     -4.822      0.001      -1.518      -0.549
POP           -0.0511      0.226     -0.226      0.826      -0.563       0.466
YEAR        1829.1515    455.478      4.016      0.003     798.788    2859.515
==============================================================================
Omnibus:                        0.749   Durbin-Watson:                   2.559
Prob(Omnibus):                  0.688   Jarque-Bera (JB):                0.684
Skew:                           0.420   Prob(JB):                        0.710
Kurtosis:                       2.434   Cond. No.                     4.86e+09
==============================================================================
```

36

A/B Testing



37

A/B Testing

- This method of introducing changes to a user experience also allows the experience to be optimized for a desired outcome and can make crucial steps in a marketing campaign more effective.
- A/B testing can also be used by product developers and designers to demonstrate the impact of new features or changes to a user experience. Product onboarding, user engagement, modals, and in-product experiences can all be optimized with A/B testing, so long as the goals are clearly defined, and you have a clear hypothesis.
- By testing ad copy, marketers can learn which version attracts more clicks. By testing the subsequent landing page, they can learn which layout converts visitors to customers best. The overall spend on a marketing campaign can actually be decreased if the elements of each step work as efficiently as possible to acquire new customers.

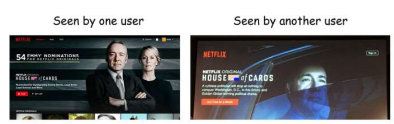
38

A/B Testing Process

- The following is an A/B testing framework you can use to start running tests:
- **Collect Data:** Your analytics will often provide insight into where you can begin optimizing. It helps to begin with high traffic areas of your site or app, as that will allow you to gather data faster. Look for pages with low conversion rates or high drop-off rates that can be improved.
- **Identify Goals:** Your conversion goals are the metrics that you are using to determine whether or not the variation is more successful than the original version. Goals can be anything from clicking a button or link to product purchases and e-mail signups.
- **Generate Hypothesis:** Once you've identified a goal you can begin generating A/B testing ideas and hypotheses for why you think they will be better than the current version. Once you have a list of ideas, prioritize them in terms of expected impact and difficulty of implementation.
- **Create Variations:** Using your A/B testing software (like Optimizely), make the desired changes to an element of your website or mobile app experience. This might be changing the color of a button, swapping the order of elements on the page, hiding navigation elements, or something entirely custom. Many leading A/B testing tools have a visual editor that will make these changes easy. Make sure to QA your experiment to make sure it works as expected.
- **Run Experiment:** Kick off your experiment and wait for visitors to participate! At this point, visitors to your site or app will be randomly assigned to either the control or variation of your experience. Their interaction with each experience is measured, counted, and compared to determine how each performs.
- **Analyze Results:** Once your experiment is complete, it's time to analyze the results. Your A/B testing software will present the data from the experiment and show you the difference between how the two versions of your page performed, and whether there is a statistically significant difference.

39

A/B Testing



Hypothesis: Proportions of Netflix users responding favorably to these two versions is different

TV series on Netflix: Users are subjected to a form of hypothesis testing called A/B testing.
Netflix shows the same show, differently designed, to different user groups. Responses of the users (click/no-click/browse/no-browse/comes-back-to-watch-or-not) are recorded and analyzed using the good old hypothesis testing method.

40

Hypothesis Testing in Python

We can perform Hypothesis testing in Python (using functions from the **Statsmodels package**).

These four situations appear in a large fraction of statistical analyses,

- One Population Proportion
- A difference in Population Proportions
- One Population Mean
- A difference in Population Means

41

42

Go to the coding demo…

43

To be continued in the next session…..

44