Top Instagram Influencers

מגישות –

מריסול קרואני

שורוק הייב

נבדוק את השפעה האינפלואנסרס באינסטגרם , נערכו כמה בדיקות עבור כמות הפוסטים ,ליקים ,מספר העוקבים וגם בדיקה ממוצע הליקים והפוסטים תוך 60 ימים .

**1) רווח מסך לתוחלת:**

נחשב את רווח הסמך ברמת סמך של 0.93 לתוחלת השפעה האינפלואנסרס זה מחושב על בסיס אזכורים, חשיבות פוסטים פופולריים

איננה ידועה מראש

נגדיר כי Y משתנה מקרי המציין את אחוז ההשפעה אחרי כל בדיקה.

בשביל גודל המדגם נניח כי

*תוצאות הבדיקות בקובץ EXCEL המצורף ->*

*נקבל מהנתונים כי :*

כעט נחשב את רווח סמך לתוחלת ברמת סמך של 0.93 :

* קודם נחשב את ערכך הביטוי

↓

* נחשב את ערך

לכן:

*לסיכום : רווח סמך של תוחלת ברמת סמך 0.93 הוא :*

מסקנה : ניתן לומר כי בסיכוי של 0.93 תוחלת של אחוז ההשפעה נעה בין לבין ובסיכוי 0.07 לא בתחום.

**2)מבחן השערה בלתי תלויים:**

החוקרים באתרי רשתות חברתיות טוענים כי ממוצע הליקים של המשפיעים מקבלים באמריקה יותר ממוצע הליקים של המשפיעים שמקבלים בהודו.

ברמת מובהקות 5%

לבדיקת הטענה ניקח 83 בדיקה שנערכו למשפיעים מאמריקה ו 25 בדיקה שנערכו למשפיעי מהודו.

|  |  |
| --- | --- |
| **India = 2** | **United state = 1** |
| 3500 K | 8300 K |
| 1600 K | 6200 K |
| 1500 K | 1900 K |
| 2900 K | 3500 K |
| 1800 K | 3700 K |
| 1600 K | 3600 K |
| 1700 K | 2400 K |
| 1500 K | 5500 K |
| 1100 K | 302 K |
| 1900 K | 329 K |
| 3100 K | 2400 K |
| 1400 K | 1700 K |
| 764 K | 2100 K |
| 897 K | 1800 K |
| 1300 K | 1300 K |
| 1100 K | 715 K |
| 1600 K | 522 K |
| 736 K | 5800 K |
| 4100 K | 3100 K |
| 695 K | 1100 K |
| 506 K | 3700 K |
| 1300 K | 2100 K |
| 1100 K | 421 K |
| 800 K | 463 K |
| 265 K | 8500 K |
|  | 1400 K |
|  | 1400 K |
|  | 1200 K |
|  | 1600 K |
|  | 204 K |
|  | 2500 K |
|  | 1200 K |
|  | 147 K |
|  | 371 K |
|  | 1900 K |
|  | 600 K |
|  | 1900 K |
|  | 1400 K |
|  | 856 K |
|  | 673 K |
|  | 4000 K |
|  | 396 K |
|  | 2300 K |
|  | 3000 K |
|  | 1400 K |
|  | 1600 K |
|  | 186 K |
|  | 4700 K |
|  | 701 K |
|  | 357 K |
|  | 2400 K |
|  | 4200 K |
|  | 1600 K |
|  | 4700 K |
|  | 2000 K |
|  | 634 K |
|  | 157 K |
|  | 4700 K |
|  | 3000 K |
|  | 1100 K |
|  | 190 K |
|  | 431 K |
|  | 1100 K |
|  | 970 K |
|  | 790 K |
|  | 3700 K |
|  | 4600 K |
|  | 3100 K |
|  | 716 K |
|  | 157 K |
|  | 623 K |
|  | 399 K |
|  | 814 K |
|  | 594 K |
|  | 1100 K |
|  | 2200 K |
|  | 1100 K |
|  | 3100 K |
|  | 136 K |
|  | 185 K |
|  | 732 K |
|  | 624 K |
|  | 720 K |

הבדיקות נערכו בשני מדינות שונים כלומר מדובר בשתי קבוצות בלתי תלויות

נניח ממוצע הליקים הן משתנים מקריים מתפלגים נורמליים

שתי השוניות אינן ידוע מראש נניח כי הן שוות

(אשר 1 = United States, 2=India)

לכן המודל המתאים להסקה על הפרש שתי התוחלות מודל t ←

שלב 1: ננסח השערה: (חד-צדדית ימנית)

שלב 2: נתונים

כאשר כ-אמדים של התחולות :

אמדים של השונות :

* n1= 83, n2= 25
* α= 0.05

כעת נחשב את :

(אומדן לסטיית התקן של )→

שלב 3 – כלל החלטה :

דחה בר"מ α= 0.05 אם Pvalue<= 0.05

שלב 4 – חישוב ערך Pvalue :

Pvalue=1-T. DIST (1.000349787,106,1) = 0.159709703

סיכום :

קיבלנו כי Pvalue גדולה מ- α= 0.05 לכן לא נדחה H0 בר"מ 0.05 נקבע כי תוחלת ממוצע הליקים באמריקה יותר מ ממוצע כמות הליקים הודו.

**3)מבחן השערה למדגמים תלויים:**

*נבחן את כמות העוקבים למשפיעים מ צרפת אשר נערכו שתי בדיקות לכמות העוקבים אחת לפני מספר חודשים והשנייה אחרי המונדיאל אשר טוענים החוקרים כי עכשווי כמות העוקבים יותר ברמת מובהקת %3(תוצאות הבדיקה בקובץ Execl מדגמים תלויים).*

|  |  |  |  |
| --- | --- | --- | --- |
| **Di = Yi - Xi** | **After(Y)** | **before(X)** | **Name** |
| 14900 K | 87600 K | 72700 K | k.mbappe |
| 3200 K | 65100 K | 61900 K | psg |
| 1000 K | 56200 K | 55200 K | paulpogba |
| 1500 K | 50200 K | 48700 K | louisvuitton |
| 800 K | 42300 K | 41500 K | dior |
| 1400 K | 37900 K | 36500 K | antogriezmann |
| 1500 K | 35000 K | 33500 K | zidane |
| 6500 K | 63400 K | 56900 K | karimbenzema |

*מדובר במקרה שלנו בשתי בדיקות אחת לפני מספר חודשים והשנייה אחרי המונדיאל כלומר התכונה שנחקרת כמות העוקבים למשפעים שרובם שחקני כדור רגל מצרפת וזה משתנה רציף ובנוסף על כך השוניות אינן ידועות מראש אזי נניח כי רמת העכירות היא נורמאלית. מסקנה המודל המתאים הסקה על הפרש התוחלת במדגמים תלויים.*

*שלב 1 : ניסוח השערה*

*שלב 2: נתונים*

* n=8,
* *נחשב את טעות התקן :*

*שלב 3 : כלל החלטה*

*נדחה את H0 בר"מ α = 0.15 אם מתקיים*  Pvalue<α

*שלב 4 : חישוב ערך* Pvalue

Pvalue =1- T.DIST(,7,1)= 0.029616294

קיבלנו כי Pvalue קטנה מ-.α= 0.03 לכן נדחה H0 בר"מ 0.03 נקבע כי תוחלת כמות העוקבים אחרי המונדיאל יותר מלפני .

**4) ANOVA חד כיוני:**

מעוניינים להשוות האם קיים הבדל המספר הפוסטים בין 3 מדינות קנדה, קולומביה, ספרד.

לצורך כך לקחנו 5 משתמשים מכל מדינה ובדקנו לפי מספר הפוסטים לכל אחד מהם להלן התוצאות:

|  |  |  |
| --- | --- | --- |
| **Canada=3** | **Colombia=2** | **Spain=1** |
| 7400 | 2000 | 6900 |
| 2500 | 8200 | 10000 |
| 680 | 10000 | 2200 |
| 670 | 1100 | 8900 |
| 1400 | 10000 | 2000 |

* מערכת ההשערות הנבחנת :

µ1-התוחלת של הטיפול ב מדינה ספרד.

µ2-התוחלת של הטיפול ב מדינה קולומביה.

µ3-התוחלת של הטיפול ב מדינה קנדה.

* טבלה ניתוח השונות בעזרת אקסל :ANOVA

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 43369000 | 2 | 21684500 | 1.586199 | 0.244772 | 3.885294 |
| Within Groups | 1.64E+08 | 12 | 13670733 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 2.07E+08 | 14 |  |  |  |  |

* מבחן F :

אם  נכונה מתקיים,

כלל החלטה: נדחה את  בר"מ  אם *נבדוק עבור* 5%=*,*כלומר קיימת לפחות תוחלת אחת ששונה משאר התוחלות*.*

* **חישוב pv:**
* *ניתן לראות כי Pv >0.05 ולכן לא נדחה את השערת האפס וניתן לומר כי בר"מ של 5% אין הבדל בין התוחלות המדינה אינו משפיע על מס הפוסטים .*

**5) ANOVA דו כיוני:**

*מנהלי חברת meta המפעילה בין השאר את הרשתות החברתיות פייסבוק ואינסטגרם...*

*רוצים לבדוק את השפעתם של הגורמים:*

* ***A-שיעור השתתפות של 60 הימים האחרונים של משתמשים.***

***B- ארץ או אזור המוצא של המשתמש.***

***על סך כל הלייקים שהמשתמש קיבל על הפוסטים שלו.***

* ***לכן לקחנו לבדוק 5 נתונים לשלוש שיעורים שונים לשלוש מדינות שונות כאשר לבדוק לכל מדינה 15 בדיקה .***

|  |  |  |  |
| --- | --- | --- | --- |
| **60\_day\_eng\_rate**  **Country** | **Brazil** | **India** | **United States** |
| 0% | 1200000 K | 439400 K | 1400000 K |
| 0% | 2300000 K | 3500000 K | 784000 K |
| 0% | 149200 K | 3500000 K | 6700000 K |
| 0% | 1700000 K | 1300000 K | 1700000 K |
| 0% | 2600000 K | 672100 K | 1900000 K |
| 1% | 5400000 K | 4900000 K | 1600000 K |
| 1% | 1700000 K | 2900000 K | 13500000 K |
| 1% | 3000000 K | 2800000 K | 2500000 K |
| 1% | 4400000 K | 3000000 K | 5300000 K |
| 1% | 1900000 K | 2200000 K | 2700000 K |
| 2% | 820100 K | 3700000 K | 4200000 K |
| 2% | 2300000 K | 810000 K | 1600000 K |
| 2% | 1900000 K | 2200000 K | 57400000 K |
| 2% | 14100000 K | 419000 K | 283000 K |
| 2% | 1500000 K | 1700000 K | 3400000 K |

*להלן טבלה הנתונים:*

**מסקנות מהטבלה באקסל נניח ש α=5%=0.05:**

* ניתן לראות כי עבור גורם A – שיעור השתתפות של 60 הימים האחרונים של המשתמש.

לא נדחה את השערת האפס בכל ר"מ סבירה, כלומר, לגורם שיעור השתתפות של 60 הימים האחרונים של המשתמש אין השפעה על סכום הלייקים.

* ניתן לראות כי עבור גורם B –- ארץ או אזור המוצא של המשתמש.

לא נדחה את השערת האפס בכל ר"מ סבירה, כלומר, לגורם ארץ או אזור המוצא של המשתמש אין השפעה על סכום הלייקים.

* ניתן לראות כי עבור גורם AB קיימת אינטראקציה בין שיעור השתתפות ל ארץ או אזור המוצא של המשתמש .

לכן נחליט כי אין לדחות את השערת האפס בר"מ של 5% כלומר האינטראקציה בין שיעור השתתפות של 60 הימים האחרונים של המשתמש ל ארץ או אזור המוצא של המשתמש על סכום הלייקים .