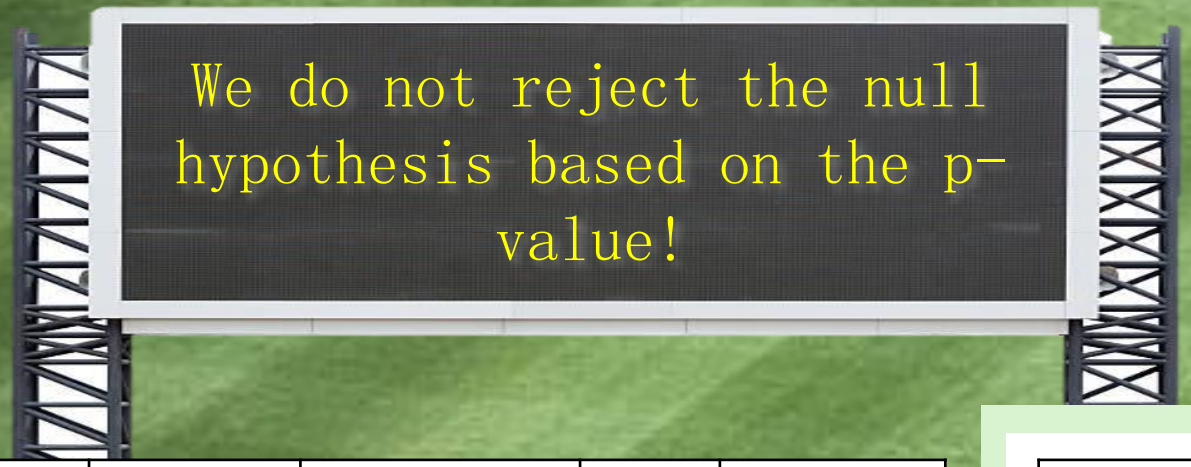

WORLD SERIES
2024


Dodgers

LAB 1: Hypothesis Testing



Mean	Z- Test value	Value from Z-Table	P-value	Double p-value
4	1.770978158	0.9616	0.0384	0.0768

Our Example:

As the Yankees prepare for the 2024 World Series against the LA Dodgers, the coaching staff is focusing on analyzing player rest days throughout the regular season, so that the team enters the postseason in peak physical condition. We have a 95% confidence interval. The double P-value is 0.0768 which is greater than alpha of 0.05, so we do not reject the null hypothesis. The Z value of 1.77 is also < 1.96, so the CV tell us to not reject the null hypothesis.

Mean	4.25
Standard Error	0.141164926
Median	4
Mode	4
Standard Deviation	0.691564075
Sample Variance	0.47826087
Kurtosis	-0.83531448
Skewness	0.268883853
Range	2.5
Minimum	3
Maximum	5.5
Sum	102
Count	24
Confidence Level(95.0%)	0.292021898

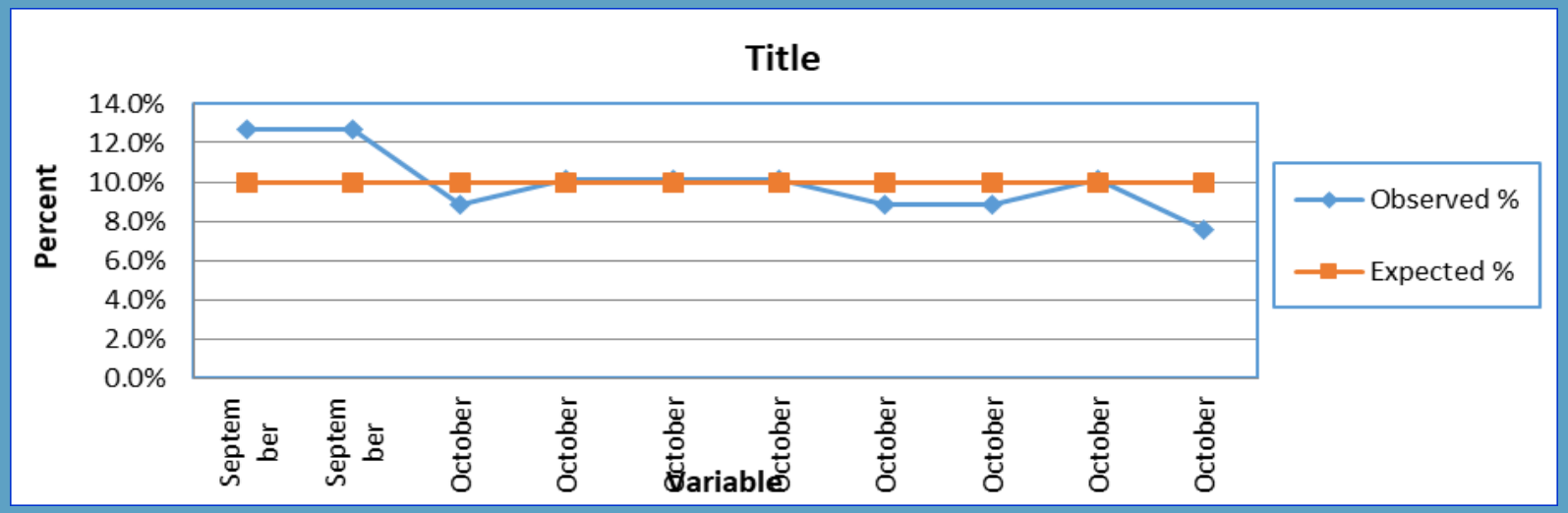
Rest Days	Month
4.0	August
3.5	August
3.5	August
4.5	August
4.5	August
4.0	August
4.0	August
5.5	September
5.0	September
4.5	September
5.0	September
4.0	September
5.0	September
5.5	September
5.0	September
5.0	September
3.5	October
4.0	October
4.0	October
4.0	October
3.5	October
3.5	October
4.0	October
3.0	October

LAB 2: Chi-Square

One-Way Chi-Square Test (Uniform Distribution/Multinomial Distribution)

Uniform	Categorical Variable Name										
	September	September	October	October	October	October	October	October	October	October	Total
Observed	5.0	5.0	3.5	4.0	4.0	4.0	3.5	3.5	4.0	3	39.5
Observed %	12.66%	12.66%	8.86%	10.13%	10.13%	10.13%	8.86%	8.86%	10.13%	7.59%	100.00%
Expected	3.95	3.95	3.95	3.95	3.95	3.95	3.95	3.95	3.95	3.95	39.5
Expected %	10.00%	10.00%	10.00%	10.00%	10.00%	10.00%	10.00%	10.00%	10.00%	10.00%	100.00%
χ^2 - Value	0.279113924	0.279113924	0.051265823	0.000632911	0.000632911	0.000632911	0.051265823	0.051265823	0.000632911	0.228481013	0.943037975
											Total

Overall χ^2 - Value	0.943038		16.9190	← χ^2 - Critical Value	α =	0.05
P-Value	0.99956		9	Degrees of Freedom	P-Value ≤ α → Significant	
Significant?	No				Expected Value < 5 → Violation	



Our Example:
The Dodgers and Yankees are preparing for the 2024 World Series. In tough games, coaches are observing how players perform. They determine whether players play even better under pressure or hit as expected with a 95% confidence range.

The hypothesis is not rejected because the p-value is below the significance level, and the chi-squared value (0.9430) is less than the critical value (16.919).

LAB 3: Simple Regression

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.681161638							
R Square	0.463981176							
Adjusted R Square	0.387407059							
Standard Error	3.261179866							
Observations	9							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	64.44183007	64.44183007	6.059242871	0.043366497			
Residual	7	74.44705882	10.63529412					
Total	8	138.8888889						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	11.51764706	3.270757507	3.521400481	0.009709356	3.783534536	19.25175958	3.783534536	19.25175958
Online Ads	1.847058824	0.750363234	2.461552939	0.043366497	0.072731723	3.621385924	0.072731723	3.621385924

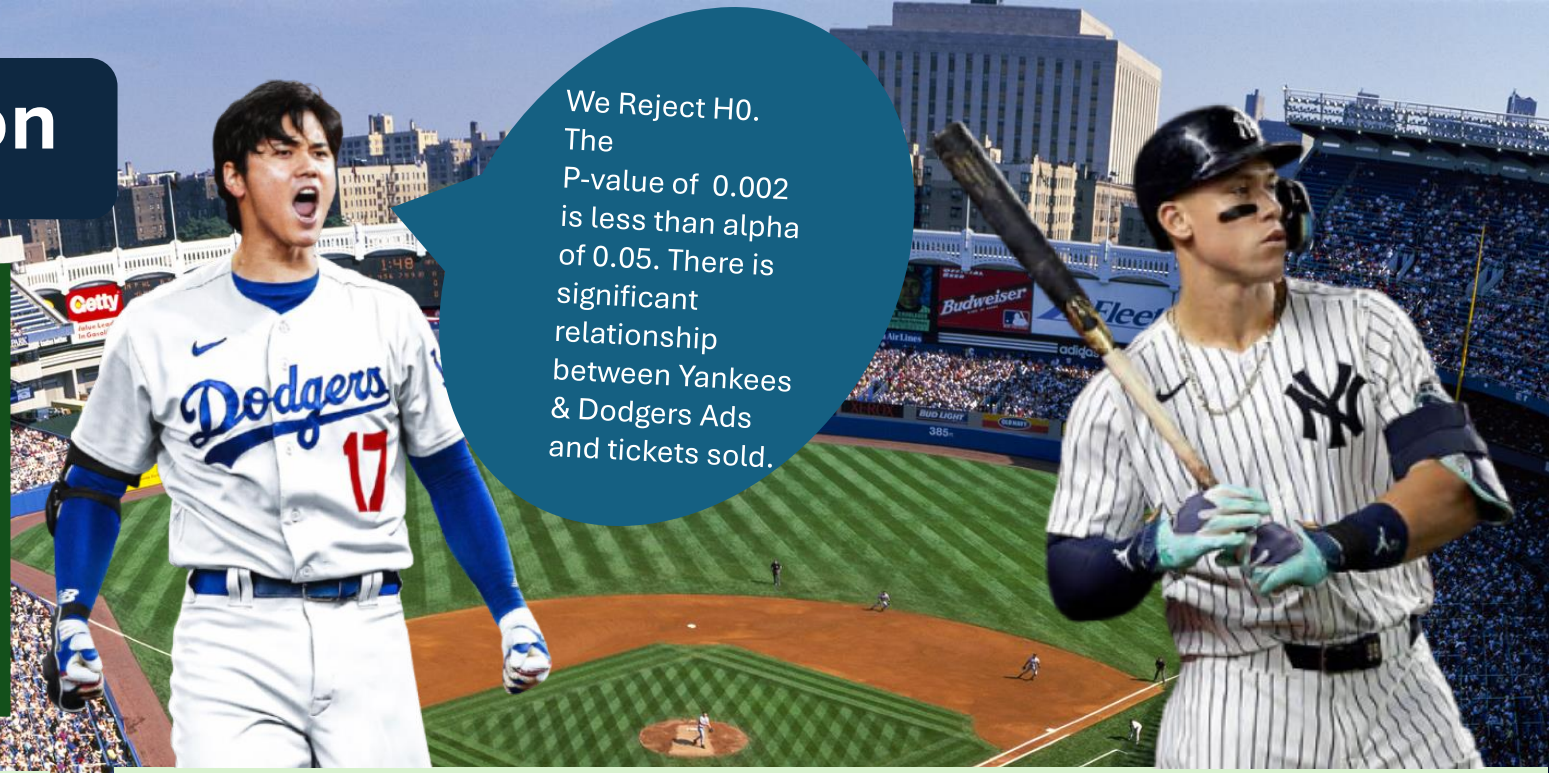
We reject H0 as the P-Value is 0.0434 as there is a significant linear relationship between the number of strikeouts and the total amount of pitches pitched.

Our Example:

The Dodgers and Yankees are preparing for the 2024 World Series and using simple regression we will model the pitchers strikeout percentage. The Dodgers believe there is no significant linear relationship between Total pitches and the strikeouts. The CI is 95%

Strikeouts	Total Pitches
5	20
4	15
6	20
2	18
6	25
5	23
4	19
2	11
5	21

Lab 4: Multiple Regression



We Reject H0. The P-value of 0.002 is less than alpha of 0.05. There is significant relationship between Yankees & Dodgers Ads and tickets sold.

Our Example:

MLB wants to analyze how ad campaigns from the Yankees and Dodgers impact ticket sales. The confidence level is 95%.

Yankees Ads	Dodgers Ads	Tickets Sold
5	6	20
4	15	15
6	6	20
2	7	18
6	4	25
5	5	23
4	8	19
2	12	11
5	6	21

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.930990256
R Square	0.866742857
Adjusted R Square	0.82232381
Standard Error	1.756318902
Observations	9

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	120.3809524	60.19047619	19.51286449	0.002366309
Residual	6	18.50793651	3.084656085		
Total	8	138.8888889			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	19.61904762	3.39497304	5.778852258	0.001173523	11.31184785	27.92624738	11.31184785	27.92624738
Yankees Ads	1.174603175	0.49971017	2.350568878	0.057014093	-0.048143564	2.397349913	-0.048143564	2.397349913
Dodgers Ads	-0.73015873	0.20992026	-3.478267084	0.013170299	-1.243815102	-0.216502359	-1.243815102	-0.216502359

LAB 5: ANOVA

Video Ads	Search Engine Ads	Social Media Ads
5	1	10
4	2	5
3	3	10
2	4	5
1	5	5

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Video Ads	5	15	3	2.5
Search Engine Ads	5	15	3	2.5
Social Media Ads	5	35	7	7.5

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	53.3333333	2	26.6666667	6.4	0.01283445	3.88529383
Within Groups	50	12	4.16666667			
Total	103.333333	14				

Our Example:
The Dodgers launched three online ad campaigns during the World Series – Video, Search Engine, and Social Media Ads. This campaign was to promote ticket sales and merchandise. The confidence interval is 95%.

We Reject Null Hypothesis! The P-Value is less than our alpha (0.01 < 0.05). There is a significant difference in at least 1 mean.

Conclusion and Credits

- Safa Alasady
- Andy Nguyen
- Yasmin Quezada
- Mia Baca
- Jonathan Nguyen
- Kelli Davis
- Milagros Monroy
- Daniel Medrano
- Maria Vaca
- Damian Flores
- Logan Demoss
- Rosa Martinez