Identifying Trends In Engine Displacement of Motorcycles

An Analysis of What Makes Motorcycles Scream

Presented by Matthew Miller

Background

Motorcycle riding is an ever evolving, beloved past time of adrenaline junkies everywhere, with new riders coming into the fold each day. According to the Motorcycle Industry Council, 1 in 4 motorcycle owners in the millennial age group is female, up from 1 in 5 for Generation X and 1 in 7 for Baby Boomers. In addition to more inclusive demographics resulting in higher demand for smaller, lightweight bikes, new technological advancements such as fuel injection permit engines to be downsized without sacrificing power. We wish to see what trends present themselves in looking at data from the past 35 years, with the objective of seeing the effects of modernization in the production of motorcycles and using it to predict the future of manufacturing within the field.

Methods

Using data from bikez.com, we will attain the engine size of every power sports vehicle with a model year from 1985 to 2020, inclusive. We will then clean the data so only motorcycles and scooters are left within the dataset and are then further divided into dirt bikes, street motorcycles, and hybrid motorcycles. Year by year averages for each category were then calculated, compared, graphed and analyzed using univariate analysis and chi square testing.

Line graphs were utilized to visualize any upward or downward trends for engine displacement both overall and within subcategories. Chi squared testing was run on each subcategory as well as the overall data to test for the goodness-of-fit of a normal distribution within the category. A p-value was calculated, which demonstrates the likeliness of each category following a normal distribution. Finally, a 95% confidence interval will be calculated at each level in order to aid with predictions in future trends as well as for the purpose of identifying outliers within the past 35 model years.

Conclusions

Table 1 contains all the points displayed on the visualization in Figure 1, as well as the number of observations for each step, as well as the overall averages of each category and the number of bikes in each. In analyzing Figure 1, we see there is no clear upward or downward trend for any category, it is more of an oscillation pattern over time. This lends itself to the idea that each category follows a normal distribution meaning future data points may be predicted with certainty to fall within a certain range.

Figure 2 provides histograms of each category as an aid in visualizing the distribution of yearly averages in each category. As the shape of the histogram agrees with the idea of a normal distribution, there is further evidence to suggest the validity of chi square testing for normalcy. Chi square values were obtained by calculating the sample variance of each category, then ranking each observation by percentile and running a chi squared test on the proportion of observations within an ideal normal curve divide into six partitions. The p-value from the overall set of 0.8014 suggests normalcy, and the evidence is only strengthened when the data is divided into subcategories.

As the datasets have been proven to be normal, agreeing with the Central Limit Theorem, we may now use the previously obtained sample variance and standard deviation to obtain a 95% confidence interval for each subcategory; that is, for any future year, there is a 95% chance the average displacement for the category falls within the calculated interval.

Table 1: Average Engine Size by Year and Category

Year	Dirt	nDirt	Hyb	nHyb	Street	nStreet	Overall	nOverall
1985	202.333	15	514.319	31	645.340	209	603.353	255
1986	215.916	22	528.103	29	675.370	232	624.562	283
1987	224.842	25	530.542	35	660.814	220	605.604	280
1988	188.940	19	565.562	32	708.615	244	659.627	295
1989	214.505	12	548.000	66	727.908	251	673.091	329
1990	237.440	16	51\70.471	51	718.442	208	663.014	275
1991	172.916	11	592.698	53	760.173	208	703.790	272
1992	245.717	7	586.283	53	770.434	199	718.569	259
1993	116.338	3	694.684	19	889.860	107	843.124	129
1994	115.500	2	739.900	22	848.176	108	819.021	132
1995	162.120	6	706.760	25	845.824	136	800.442	167
1996	200.670	3	641.000	36	842.510	161	796.611	200
1997	178.000	21	633.591	44	756.019	211	692.522	276
1998	162.976	32	619.060	50	761.885	211	672.102	293
1999	176.932	26	592.618	37	807.733	189	711.065	252
2000	231.893	29	567.034	54	831.411	241	733.687	324
2001	169.768	37	610.100	60	819.769	270	719.958	367
2002	156.393	100	533.462	49	812.773	263	620.238	412
2003	133.131	85	590.039	65	855.191	293	677.741	443
2004	171.486	168	509.738	85	704.092	336	524.127	589
2005	133.770	269	500.280	86	655.540	388	448.665	743
2006	119.462	408	521.391	134	697.245	663	482.058	1205
2007	118.383	540	468.234	154	618.401	846	428.051	1540
2008	127.596	507	493.384	138	625.889	962	457.299	1607
2009	128.669	430	523.858	115	857.269	953	622.528	1498
2010	136.012	549	422.988	156	773.390	1064	644.683	1769
2011	148.556	426	519.437	139	774.170	1005	581.862	1570
2012	161.763	381	591.134	125	722.714	913	560.509	1419
2013	147.311	317	595.332	113	767.087	828	595.483	1258
2014	176.311	272	576.367	110	810.315	778	639.466	1160
2015	155.056	212	569.655	127	813.578	817	666.012	1156
2016	151.640	430	688.412	145	730.685	942	562.512	1517
2017	169.183	311	650.504	132	765.419	801	604.164	1244
2018	186.866	200	607.770	153	839.512	710	683.362	1063
2019	195.289	265	694.419	135	762.511	707	618.420	1107
2020	170.866	123	643.423	62	812.684	392	657.678	577
Total	**148.883**	**6279**	**569.098**	**2807**	**753.635**	**17066**	**588.629**	**26152**

Figure 1: Average Engine Size Over Time

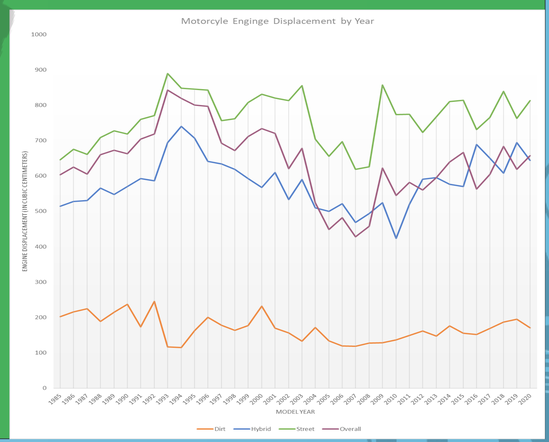
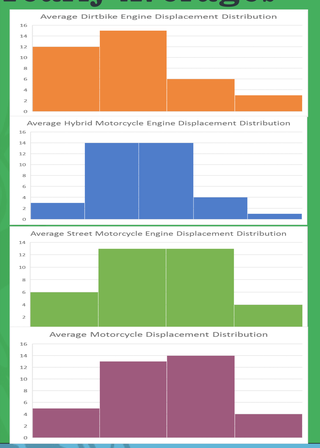


Figure 2: Distribution of Yearly Averages



X^2=1.667
P=0.8931
95% Confidence Interval: [97.847 , 241.259]

X^2=1.500
P=0.9131
95% Confidence Interval: [438.828 , 724.536]

X^2=0.667
P=0.9847
95% Confidence Interval: [619.737 , 906.305]

X^2=2.333
P=0.8014
95% Confidence Interval: [443.888, 820.584]