Stats and Probability Project 2025
Lauren H Brampton

Excel Probability

I used excel to calculate the probabilities for binomial, normal and poission distributions. I used the built in functions which include BINOM.DISt, NORM.DIST and POISSION.DIST to find the exact values for each scenario that was given to us. I also used BINOM.DIST.RANGE and the normal approximation to estimate the probability of getting between 120 and 150 heads in 250 coin tosses. Everything is layed out in the table with scenarios, formulas and results.

	A	B	C	D
1	Distribution	Scenario	Formula	Results
2	Binomial	15 heads in 28 tosses of a fair coin	P(X = 15) p = 0.5	0.139482915
3	Binomial	9 rolls of a 5 in 50 rolls of a fair dice	P(X = 9) p = 1/6	0.140958274
4	Binomial	9 totals of 6 in 45 rolls of 2 fair dice	P(X = 9), p = 5/36	0.078280242
5	Normal	More than 36 minutes	P(X > 36), mean = 29, sd = 4	0.040059157
6	Normal	Less than 21 minutes	P(X < 21), mean = 29, sd = 4	0.022750132
7	Normal	Between 22 and 36 minutes	P(22 < X < 36), mean = 29, sd = 4	0.919881686
8	Poisson	More than 3 calls in a minute	P(X > 3), lambda = 5.4	0.786708982
9	Poisson	4 calls	P(X = 4), lambda = 5.4	0.160019753
10	Poisson	Less than 6 calls	P(X < 6) = P(X <= 5), lambda = 5.4	0.546132104
11	Poisson	Between 2 and 6 calls	P(2 <= X <= 6) = P(X <= 6) - P(X <= 1), lambda	0.672765186
12	Binomial	Between 120 and 150 heads (250 tosses)	P(120 <= X <= 150), p = 0.5	0.756047538
13	Normal Approx to Binomial	Between 120 and 150 heads (approx)	P(119.5 < X < 150.5), mean = 125, sd = 7.91	0.755940772
14				

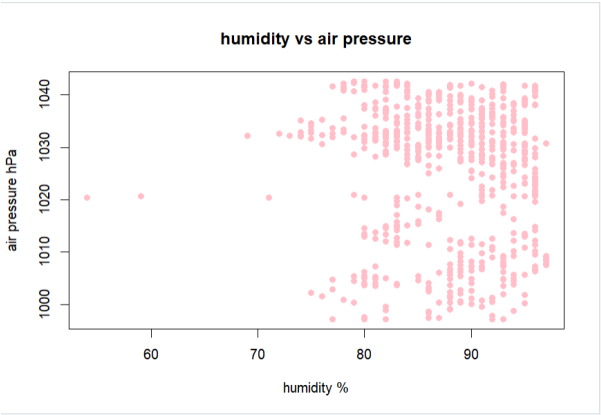
R Statistical Analysis

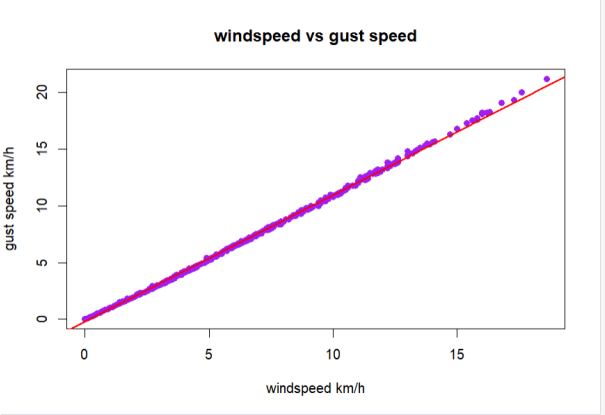
In this part i used R to load the weather data which was on moodle. I did mine on the humidity column and calculated 10 statistical quantities. I had to convert the data to numeric at first since it was read in as text. The average humidity was around 88%, most values were between 84% and 93%. These values gave me a good overview of how the data was set out.

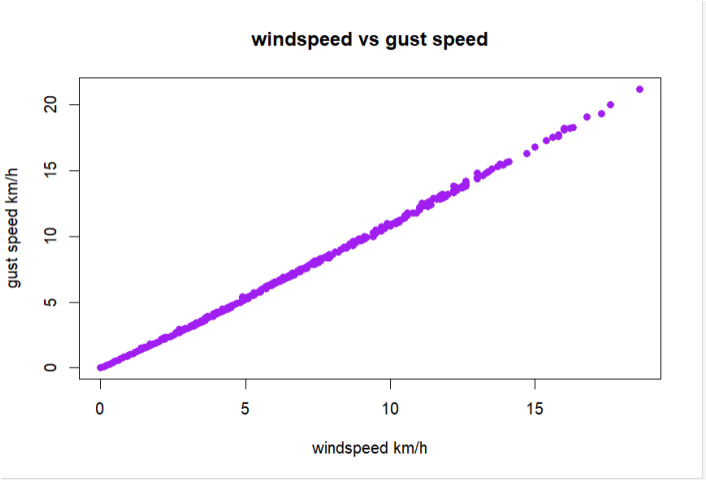
```
> #calculating stats for humidity
> mean(humidity, na.rm = TRUE)#ave humidity
[1] 87.96439
> median(humidity, na.rm = TRUE)#middle val
[1] 89
> min(humidity, na.rm = TRUE)#lowest humidity
[1] 54
> max(humidity, na.rm = TRUE)#highest humidity
[1] 97
> sd(humidity, na.rm = TRUE)#standard deviattion
[1] 5.844894
> var(humidity, na.rm = TRUE) #variance
[1] 34.16278
> range(humidity, na.rm = TRUE)#range
[1] 54 97
> IQR(humidity, na.rm = TRUE)#interquartile range
[1] 9
> quantile(humidity, 0.25, na.rm = TRUE)#25th percentile
25%
 84
> quantile(humidity, 0.75, na.rm = TRUE)#75th percentile
75%
 93
>
```

R Graphs and Regression

Next I used R to plot two scatterplots. One was showing humidity vs air pressure and then other showed windspeed vs gust speed. I added a linear regression line to the second plot using lm() and abline(). The relationship between windspeed and gist was very song and nearly perfectly linear.







Diamond Price Regression in Excel

I used exel to analyse the relationship between diamond carat and price. I created a scatterplot and added a trendline with y=4018.7x - 888.75, this equation let me calculate fitted prices and i worked out the residuals (price minus predicted) for each row. The standard error of estimate was 7.4494 and the coefficient was 0.9998. This shows a nearly perfect linear line between carat and price.

	A	B	C	D
1	Diamond Prices			
2	carat	price	Fitted price	Residual (price-fitted)
3	0.5	1125	1120.6	4.4
4	0.51	1167	1160.787	6.213
5	0.52	1192	1200.974	-8.974
6	0.53	1247	1241.161	5.839
7	0.54	1281	1281.348	-0.348
8	0.55	1307	1321.535	-14.535
9	0.56	1359	1361.722	-2.722
10	0.57	1401	1401.909	-0.909
11	0.58	1459	1442.096	16.904
12	0.59	1481	1482.283	-1.283
13	0.6	1519	1522.47	-3.47
14	0.61	1567	1562.657	4.343
15	0.62	1609	1602.844	6.156
16	0.63	1640	1643.031	-3.031
17	0.64	1682	1683.218	-1.218
18	0.65	1727	1723.405	3.595
19	0.66	1764	1763.592	0.408
20	0.67	1789	1803.779	-14.779
21	0.68	1843	1843.966	-0.966
22	0.69	1890	1884.153	5.847
23	0.7	1905	1924.34	-19.34
24	0.71	1969	1964.527	4.473
25	0.72	2007	2004.714	2.286
26	0.73	2051	2044.901	6.099
27	0.74	2087	2085.088	1.912
28	0.75	2130	2125.275	4.725
29	0.76	2160	2165.462	-5.462
30	0.77	2210	2205.649	4.351
31	0.78	2243	2245.836	-2.836
32	0.79	2289	2286.023	2.977
33				
34			Standard error	7.449400252
35			Correlation Coefficent	0.999786032
36				
37				

Conclusion

In conclusion this project helped and apply prob, stats and regression analysis in a practical way using excel and R. I fee quite confident using these tools to work with data. I also got more familiar with how distributions work and how to fit and interpriate trendlines and residuals.