Assessment 2

UTS
UNIVERSITY OF TECHNOLOGY SYDNEY

23708 - Studio 1: Foundation - Spring 2024

Udemy

Somesh
Shanbhag

25525837

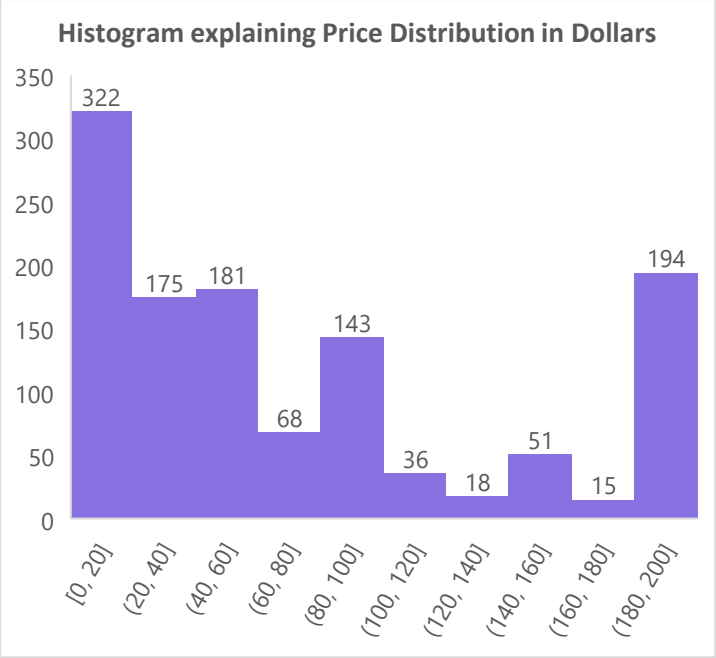
Table of Contents

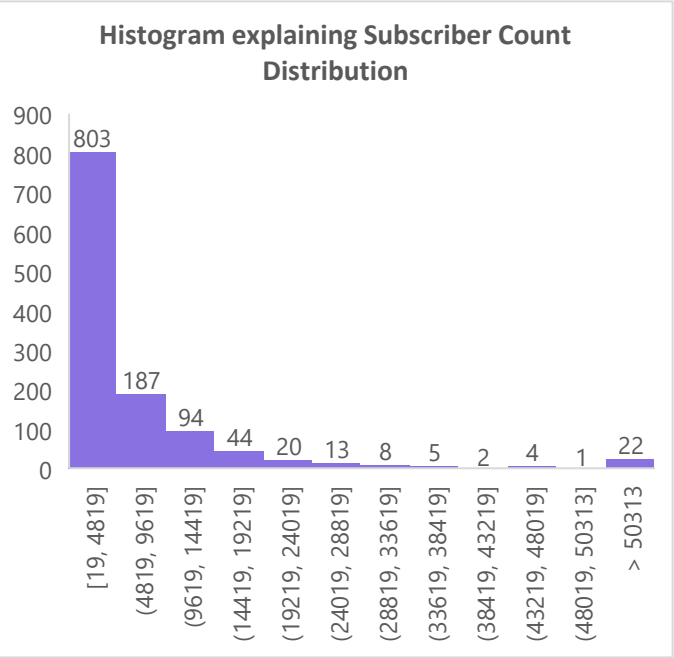
Question 1

Q. Provide summary statistics for the following variables: price; num_subscribers; num_reviews; num_lectures; rating; content_duration. Include all measures of central tendency and at least one measure of variability. Provide a graph (hint: a histogram) to visually show the distribution of the price and num_subscribers variables and comment on features of the distribution.

A. Summary Statistics are as follows:

price		*num_subscribers*		*num_reviews*		*num_lectures*		*Level_category*		*rating*		*content_duration*	
Mean	77.04	Mean	6635.02	Mean	357.36	Mean	52.77	Mean	2.71	Mean	0.64	Mean	5.59
Standard Error	1.91	Standard Error	419.77	Standard Error	45.35	Standard Error	1.81	Standard Error	0.04	Standard Error	0.01	Standard Error	0.21
Median	50.00	Median	2430.00	Median	65.00	Median	32.00	Median	4.00	Median	0.76	Median	3.00
Mode	20.00	Mode	1703.00	Mode	16.00	Mode	20.00	Mode	4.00	Mode	0.96	Mode	1.00
Standard Deviation	66.15	Standard Deviation	14559.42	Standard Deviation	1572.92	Standard Deviation	62.76	Standard Deviation	1.40	Standard Deviation	0.31	Standard Deviation	7.24
Sample Variance	4375.22	Sample Variance	211976681.83	Sample Variance	2474070.04	Sample Variance	3939.10	Sample Variance	1.96	Sample Variance	0.09	Sample Variance	52.46
Kurtosis	-0.75	Kurtosis	110.91	Kurtosis	145.17	Kurtosis	27.25	Kurtosis	-1.83	Kurtosis	-1.10	Kurtosis	22.18
Skewness	0.76	Skewness	8.44	Skewness	10.96	Skewness	4.18	Skewness	-0.25	Skewness	-0.59	Skewness	3.86
Range	200.00	Range	268904.00	Range	27445.00	Range	774.00	Range	3.00	Range	1.00	Range	76.00
Minimum	0.00	Minimum	19.00	Minimum	0.00	Minimum	5.00	Minimum	1.00	Minimum	0.00	Minimum	0.50
Maximum	200.00	Maximum	268923.00	Maximum	27445.00	Maximum	779.00	Maximum	4.00	Maximum	1.00	Maximum	76.50
Sum	92675.00	Sum	7981935.00	Sum	429899.00	Sum	63479.00	Sum	3264.00	Sum	773.59	Sum	6720.75
Count	1203.00	Count	1203.00	Count	1203.00	Count	1203.00	Count	1203.00	Count	1203.00	Count	1203.00





From the Summary Statistics table, we can understand that:-

- **Price**: We see that the distribution for Price values is right skewed from the 0.76 skewness value. This means that more courses are priced at the lower end, but there are a few very high-priced courses as well since our mean is $77, which is much higher than the median of $50. The mode highlights that a majority of courses are priced at $20, suggesting that Udemy understands its target market, which is usually students who are price-conscious and are trying to keep the majority of its courses affordable while also trying to provide value. However, since we also have a maximum price of $200, this distribution creates a long right tail. A kurtosis of -0.75 suggests a comparatively flatter distribution, indicating few extreme values compared to the normal distribution.
- **Number of Subscribers**: The subscriber count shows extreme right skewness of 8.44 and very high kurtosis of 110.91 indicating that comparatively the majority of a courses have a small subscriber count while a small number of courses have disproportionately large number of subscribers. This drives the mean subscriber count which is 6,635 significantly higher than the median which is 2,430. The mode is 1,703 suggesting that most courses cluster around this subscriber count with a few outliers pulling the distribution significantly upwards. The high standard deviation of 14,559 along with the range depicting variation from 19 to 268,923 subscribers further highlights the wide variability in subscriber numbers across courses.
- **Number of Reviews**: The number of reviews does show extreme skewness with a value of 10.96 indicating that most courses receive a small number of reviews, while a few courses have many reviews, which drives the mean of 357 far above the median of 65. The kurtosis of 145.17 suggests a very peaked distribution explaining that most courses fall with a narrow range of review counts but there are definitely some extreme outliers present as well. The high standard of deviation (1,572) and range (0 to 27,445) indicate significant variation in review counts. If Udemy gave some discount or incentive to leave reviews, we might see much higher review counts for each course.
- **Number of Lectures**: The distribution of lecture counts is right skewed with a value of 4.18 with most courses offering fewer lectures, while a few offer many more which we can see from the maximum of 779. The mean of 52.77 is much higher than the median of 32, highlighting that the effect of a few courses with high lecture counts is pulling the average upwards. The mode which is 20 and median which is 32 indicate that most courses are shorter while a few extensive courses increase the spread of the distribution.
- **Level Category**: The level category is fairly evenly distributed with a slight skewness of -0.25 suggesting a near symmetric distribution though there are more advanced courses than beginner ones. The mode of 4 suggests that most courses fall into the levels category. The relatively low SD suggests that there is very little variability in this field.
- **Ratings**: The negative skewness -0.59 indicates a slight skew towards higher ratings, with more courses rated closed to 1. This shows that most courses tend to have good reviews.

The mode of 0.96 indicates that the most frequent rating is close to the maximum of 1 and median rating is 0.76 which still is relatively high. The kurtosis of -1.10 suggests a flatter-than-normal distribution which explains that ratings are more spread out than in a normal distribution but with a bias towards higher ratings.

- **Content Duration**: The content duration shows a strong right skewness of 3.86 meaning most courses are short but there are a few courses that are much longer that pull the mean (5.59) higher than the median (3). The mode which is 1 hour reflects that the most common course length is short while some of the courses have a duration up to 76.5 hours. The quite high kurtosis of 22.18 indicates that while most courses are clustered around a short duration, there are definitely some extreme outliers with significantly longer durations.

Our observations from the graphs are similar, On the topic of central tendency and statistical insight, Price Distribution is moderately skewed, with a peak at the lower price range as explained earlier. The average price is higher than the median, showing that the right tail of higher-priced courses pulls the mean upwards. Subscriber distribution on the other hand is highly right skewed with significant outliers. The median is much lower than the mean, indicating that a few popular courses have many more subscribers than the rest.

On the topic of variability, Price is moderately variable indicating a wide range of pricing for courses as opposed to subscribers which is highly variable driven by a few very popular courses.

Question 2

Q. Use an appropriate chart or table to summarise the frequencies for the following variables: **Level_category** and **year**. Please note that the variable **Level_category** contains information on the difficulty level of the course, while the variable **level_category value label** explains the meaning of the course level categories, e.g. level_category=1 means that all courses in category 1 are Beginner level. Briefly discuss the results.

A. Tables that summarise the frequencies for the Level Category and Year are as follows:-

Level Category	Frequency	% of Total
All Level (1)	422	35.08%
Beginner Level(2)	134	11.14%
Intermediate Level (3)	14	1.16%
Expert Level (4)	633	52.62%
Other	0	0.00%
	1203	100%

Year	Frequency	% of Total
2011	5	0.42%
2012	19	1.58%
2013	55	4.57%
2014	115	9.56%
2015	336	27.93%
2016	449	37.32%
2017	224	18.62%
Other	0	0.00%
	1203	100%

Level Category: Level Category 1 (Beginner level) is a category that has 422 courses which make up 35.08% of the total courses. A significant portion of the courses designed for beginners reflects the understanding that creating a high supply for the high demand in introductory level courses allows Udemy to then increase the number of customers willing to buy the higher level courses based on the completion of the introductory level courses. This also shows that these courses target learners just starting their journey in a particular subject. Category 2 (intermediate level) is a category that has 134 courses constituting only 11.14% of the total. This is considerably lower than category 1, indicating that fewer courses are tailored for intermediate learners. Level Category 3(Expert level): this category has only 14 courses, accounting for 1.16% of the total. Advanced courses are rare, suggesting that few courses are created for highly skilled or specialized learners which may also reflect that the demand for this expert level course might very low and specialized to only a few interested students or there might be a gap in course offering for advanced learners. Level category 4(All levels) is the largest category with 633 courses or 52.62% of the total. Courses labelled as “all levels” suggest a broad applicability and appeal to learners at any level. The dominance of this category suggests that many courses are designed to cater to a wider audience, without requiring prerequisites which could be beneficial for maximizing reach and inclusivity.

Year: Only 5 courses were created in 2011 making up just 0.42% of the total. This likely reflects the early stages of online course creation or lower interest at the time. In 2021 there were nearly 19 courses created making up 1.58%. The small increase from the previous year indicates a slow but steady growth in the number of available courses. In 2013, 55 courses were available which is around 4.57% of the total. There is a marked increase in course offerings, signalling a growing interest in online learning. In 2014 a jump of 115 courses (9.56%) occurred this year, further highlighting the growing interest and expanding market for online education. The number of courses more than doubled in 2015 to 336 courses representing 27.93% of the total. This year marks a substantial boom in online course creation, possibly due to increased awareness, technology adoption and demand. 2016 was the year that available courses peaked with 449(37.32%) the largest proportion of the dataset. This suggests a period of rapid expansion in the number of online courses offered possibly due to a convergence of factors such as increased internet access, popularity of online learning platforms, ease of self-paced learning and a shift in how people consume educational content. In 2017, a decrease to 224 courses (18.62%) suggested that the market may have begun to stabilize or mature after the explosive growth of the previous years, leading to a market correction in the demand and hence in turn the production of educational content.

Question 3

Q. Test the hypothesis that the average course rating (variable "rating") is the same in 2014 and 2017 at the significance level of 0.05 (5% significance level). *[Hint: 1. test equal variance; 2. a two-sample t-test with equal or unequal variances.]* Explain in words what are the null and alternate hypotheses.

A. Null Hypothesis (H_0): The Average course rating in 2014 and 2017 are the same.

$H_0 : \mu_{2014} = \mu_{2017}$

Alternative Hypothesis (H_1): the average course rating in 2014 and 2017 are different.

$H_1 : \mu_{2014} \neq \mu_{2017}$

Step 1: F-Test: Test for equal Variance

	2014 Ratings	*2017 rating*
Mean	0.627478261	0.659391304
Variance	0.084404989	0.089146117
Observations	115	115
df	114	114
F	0.946816204	
P(F<=f) one-tail	0.385489131	
F Critical one-tail	0.733888543	

The F-test compares the variance of two samples to check if they are significantly different. The null for the F-test is that the variances are equal.

Since the P-Value (0.3855) is greater than 0.05, we fail to reject the null hypothesis of equal variance. This means that the variance of the ratings in 2014 and 2017 are statistically similar.

Step 2: t-Test: Two-Sample Assuming Unequal Variances

	2014 Ratings	*2017 rating*
Mean	0.627478261	0.659391304
Variance	0.084404989	0.089146117
Observations	115	115
Hypothesized Mean Difference	0	
df	228	
t Stat	-0.821492213	
P(T<=t) one-tail	0.206112063	
t Critical one-tail	1.651564228	
P(T<=t) two-tail	0.412224126	
t Critical two-tail	1.970423195	

The T-Stat value is -0.82 and the two tailed P-value is 0.4122.

The critical value for t at a 5% significance level is 1.9704.

Since the P-Value is greater than 0.05, and since we cannot support the alternative hypothesis, we fail to reject the null Hypothesis. This indicates that there is no statistically significant difference in the average course ratings between 2014 and 2017 and thus we reject the alternative hypothesis.

Question 4

Q. Do you expect a positive, negative or no correlation between the price and the number of subscribers? Make a prediction about this correlation and explain your reasoning. Then calculate the correlation coefficient between the variables **price** and **num_subscribers**. Comment on the result and if it is different from your prediction, give a possible reason(s) why.

A. Based on my intuition, I would predict a negative correlation between the price of a course and the number of subscribers. The reasoning behind this prediction is as follows:-

Price sensitivity: Usually, as the price of a product or service increases, fewer people might be willing to pay for it. This is especially true for online courses, where users often look for affordable options.

Demand Effect: High-priced courses could attract fewer subscribers because potential customers may have access to similar content at a lower cost.

Hence, I assume there will be a negative correlation between price and the number of subscribers.

Actual Correlation Table:-

	price	*num_subscribers*
price	1	
num_subscribers	0.013978844	1

The correlation coefficient between price and number of subscribers is ~ 0.014.

The value 0.014 is very close to zero, indicating that there is almost no linear relationship between the price of a course and the number of subscribers. This result shows that statistically, the price and the number of subscribers are largely uncorrelated based on this dataset.

The result of no significant correlation is quite different from our prediction of a negative correlation. The initial expectation of a negative relationship was on the assumption that higher prices would deter more subscribers.

However, the data suggests that the price and number of subscribers are not meaningfully correlated. This could be due to several reasons:

Course Quality and perceived value: Higher priced courses may offer better content, prestigious instructors or certifications which could attract more subscribers despite the higher price. If subscribers perceive high value, they may be willing pay more.

Discounts and Promotions: Online platforms frequently give discounts which means that even high-priced courses may often be sold at much lower prices, reducing the impact of the listed price on subscriber numbers.

Market Segmentation: There may be different market segments, where users prioritise various criteria such as content quality, specific course features along with price that may dilute any clear price-subscriber relationship.

Popular Topics: Courses made on trending topics or new technologies may attract a high number of subscribers simply based on subject demand.

Reputation: Courses offered by some well-known organizations or instructors may also attract high number of subscribers even if the courses isn't very high in quality simply due to the value of certification from this organisation.

Question 5

Q. Use a linear regression model to estimate the effect of price on the number of subscribers. [Note that in this model, the dependent variable (Y) is the number of subscribers, and the explanatory variable (X) is price]. Estimate of the model with and without including additional explanatory variables (so-called control variables). Your control variables should include the number of lectures (**num_lectures**), the content duration (**content_duration**), and the dummy variables for different years and subject level categories (**level_category**).

A. Summary Output for Price and Number of Subscribers Only

Regression Statistics	
Multiple R	0.013978844
R Square	0.000195408
Adjusted R Square	-0.000637069
Standard Error	14564.05594
Observations	1203

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	49789191.91	49789191.91	0.234730974	0.628126199
Residual	1201	2.54746E+11	212111725.5		
Total	1202	2.54796E+11			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	6397.99	644.73	9.92	0.00	5133.06	7662.92	5133.06	7662.92
price	3.08	6.35	0.48	0.63	-9.38	15.54	-9.38	15.54

Summary Output for Multiple Variables

Regression Statistics	
Multiple R	0.32832433
R Square	0.107796865
Adjusted R Square	0.098799859
Standard Error	13821.48384
Observations	1203

	df	*SS*	*MS*	*F*	*Significance F*
Regression	12	27466207071	2288850589	11.98141478	3.34241E-23
Residual	1190	2.2733E+11	191033415.5		
Total	1202	2.54796E+11			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	29450.39	7250.66	4.06	0.00	15224.88	43675.89	15224.88	43675.89
price	-4.43	6.68	-0.66	0.51	-17.54	8.69	-17.54	8.69
num_lectures	0.23	12.61	0.02	0.99	-24.51	24.97	-24.51	24.97
content_duration	300.73	108.61	2.77	0.01	87.65	513.81	87.65	513.81
2012	-6174.28	6954.13	-0.89	0.37	-19818.01	7469.44	-19818.01	7469.44
2013	-1641.50	6480.09	-0.25	0.80	-14355.19	11072.18	-14355.19	11072.18
2014	-12845.70	6339.82	-2.03	0.04	-25284.18	-407.22	-25284.18	-407.22
2015	-15160.56	6250.41	-2.43	0.02	-27423.62	-2897.50	-27423.62	-2897.50
2016	-17797.35	6238.69	-2.85	0.00	-30037.40	-5557.29	-30037.40	-5557.29
2017	-19433.14	6278.64	-3.10	0.00	-31751.57	-7114.70	-31751.57	-7114.70
All Level (1)	-7633.03	3765.29	-2.03	0.04	-15020.38	-245.68	-15020.38	-245.68
Beginner Level(2)	-9247.79	3890.55	-2.38	0.02	-16880.89	-1614.70	-16880.89	-1614.70
Expert Level (4)	-8671.54	3744.41	-2.32	0.02	-16017.92	-1325.17	-16017.92	-1325.17

Question 6

Q. Interpret the coefficient estimates you obtained in Q5. Comment on their size and statistical significance (comment on each coefficient). Explain why the coefficient for price changes when you include other explanatory variables.

A. From the linear regression model without control variables, below is the detailed interpretation of each coefficient in the model:-

Intercept(6397.99):- This represents the predicted number of subscribers when the price is zero. It suggests that if the course were free, it may attract approximately 6,398 subscribers. This high value reflects baseline demand irrespective of price.

Price(3.08): The coefficient for price is positive but extremely small and its P-value (0.628) is well above the typical significant threshold of 5%. This suggests that price has no statistically significant impact on the number of subscribers when no control variables are included.

From the linear regression model with control variables:-

Intercept(29450.39): This represents the predicted value when all other control variables (Price,number of lectures , etc.) are 0.It's statistically significant as P value is less that 0.05, meaning that the base level of subscribers for a course with these characteristics is around 29450.

Price(-4.43): in the multivariable regression model, the price coefficient becomes negative, but its p-value (0.51) is still far from significant. This implies that price has no statistical effect on the number of subscribers, even after controlling for other factors such as number of lectures, content duration etc. The negative coefficient suggests that as price increases, subscribers are likely to decrease by 4.43, but the lack of significance suggests this finding is not reliable.

Number of lectures(0.23): This coefficient suggests that for every additional lecture in a course, the number of subscribers may increase by 0.23, however the p value is 0.999 which shows that it has no significant effect on the number of subscribers. Hence, the number of lectures offered in a course does not significantly influence subscription counts.

Level Category: The dummy variables for Level category are all negative and significant as they all have a P value less than 0.05, meaning courses at beginner, intermediate and all levels attract fewer subscribers with respect to the reference category which is expert level, which makes sense as those subscribers specifically looking for expert level courses would research courses at their difficulty level and then join which why this criteria is more likely to be a reason for increase in subscriber count as opposed to subscribers of courses with various other levels.

Content Duration(300.73): Content duration has a positive and statistically significant effect on the number of subscribers with a P-value of 0.01. For every additional unit of content duration, the number of subscribers may increase by around 300.73 subscribers. This makes sense as courses with more content may appear more valuable to users thereby attracting more subscribers.

Year(-3234.89): The dummy variables for Year show how the number of subscribers vary over time compared to the base year which is 2011. The negative coefficients for years from 2014 onwards suggest that, on average, courses offered in these years have had fewer subscribers than the base year. Since these results are statistically significant, indicating a meaningful downward trend in subscribers over time.

The price coefficient changes from 3.08 to -4.43 in the multiple regression model. This change occurs because in the multiple regression model other factors such as content duration, etc. are controlled for as they might have been confounding variables. These control variables might have an effect on both price and the number of subscribers which could cause the coefficient on price to change. For example some courses could be priced higher because they offer more content or are at higher levels. The initial positive relationship between the price and subscribers disappears when we control for these other variables as they capture some of the variation in the number of subscribers. We now start to see that price has a clear negative yet insignificant effect on the number of subscribers.

Question 7

Q. How much variation in number of subscribers is explained by the independent (i.e. explanatory) variables in your model? What does this tell us about our ability to predict demand for web courses?

A. R Square value without control variables: R Sq. value is 0.000195 meaning that only 2% of the variation in the number of subscribers is explained by price alone. This suggests that price is not a good predictor of the number of subscribers in this simple model.

R Sq. with control variables: After adding the control variables, the Adjusted R sq. value increases to 0.098 which means that a lot of new data has been introduced by the addition of explanatory variables and from R Sq. value, approximately 10% of the variation in the number of subscribers is explained by price, content duration, number of lectures, level category and year. While this is an improvement over the original model, it still suggests that the variables included leave 90% of the variation in subscriber count unexplained. This indicates that other factors not included in the model are likely to be critical in being able to influence and hence accurately predict the demand for web courses.

Questions 8

Q. Do you think the results in Q5 tell you the causal effect of a change in price on demand for web courses? That is, does it reflect the direct impact of price on demand, or could it be picking up other factors? If you think there could be other factors, provide at least one example.

A. The results suggest that price does not have a direct statistically significant impact on demand, which is number of subscribers. The coefficient for price is not significant in either model meaning we cannot conclusively say that changes in price cause changes in demand. Instead, it is likely that other potentially confounding factors such as course content, duration, organization/instructor reputation or promotional discounts are driving demand. Factors such as reputation of organization creating the web course or instructor reputation could be significant factors influencing demand. Web courses made by popular organizations for example Google, may attract more subscribers regardless of price. While courses with lower ratings may struggle to attract subscribers even if they are priced lower.

Question 9

Q. Your manager wants to know what factors are most important for predicting demand for web courses. Based on your results in Q5, what types of courses would you recommend if you want to ensure a high number of subscribers?

A. Based on the multiple regression results, content duration and year appear to be the most important factors in predicting demand for web courses. Specifically, longer content duration is positively associated with more subscribers count, meaning that courses offering more extensive material tend to attract more users. Year has significantly negative impact suggesting that courses offered in earlier years attracted more subscribers compared to recent years. This also suggests that course novelty or perceived valyu over time might influence demand. To ensure high subscribers it would be advisable to either focus on course duration creating courses with more extensive content as this seems to attract more subscribers, or ensure that courses are refreshed and remain relevant and updated to counter the negative trend over the years or pay attention to competition because courses offered in recent years seem to struggle with subscriptions which means it is important to differentiate new courses by offering unique high quality content or deploying new marketing strategies to attract new customers.

Question 10

Q. An economist at your company points out that the relationship between prices and quantity is often not linear. She recommends that you re-run your regression with a log-log demand function*. With this specification, the price and quantity are replaced with their natural logarithms (plus a small constant value for cases where price = 0). Re-run your regressions but replace the price and quantity with their log values, which are already included in the dataset. Does your conclusion about the relationship between price and quantity change? In what way? How does the model fit change? Do you think this is a more appropriate specification (explain your reasoning).

A. In this log-log regression model, the interpretation of the coefficients change. Each coefficient now represents the percentage change in the dependent variable which is number of subscribers in response to a 1% change in the corresponding explanatory variable.

The coefficient for log price is -0.154 and it is statistically significant P value of 0.00 which tells us that for 1% increase in price, we associate a 0.154% decrease in quantity demanded. This is a meaningful change from the previous linear specification where price had no significant impact.

The log-log model provides stronger evidence of a casual relationship between price and number of subscribers. The negative and significant coefficient on price suggests that increasing the price leads to a decrease in subscribers, holding all other factors constant. However, it is still clear as with the previous model that there could still be other factors influencing demand for courses.

The model fit has definitely changed as the adjusted R Sq. value has now increased to 22.1% meaning that there definitely was significant increase in new information added to this model compared to the original model. The R. Sq. value of 0.226 means that this model can explain almost 22.6% of the variations in the number of subscribers providing a clearer and more interpretable picture about how percentage changes in price and other factors affect the number of subscribers. The negative relationship between price and subscribers is more apparent and interpretable in this model.

I think this model is definitely more appropriate because the relationship between price and demand is rarely linear in real world data, particularly for something like web based services like courses. This is clearly visible from the improved R Sq. value. This means that the log-log model is able to show a far more accurate picture of the relationship between price and subscriber count.

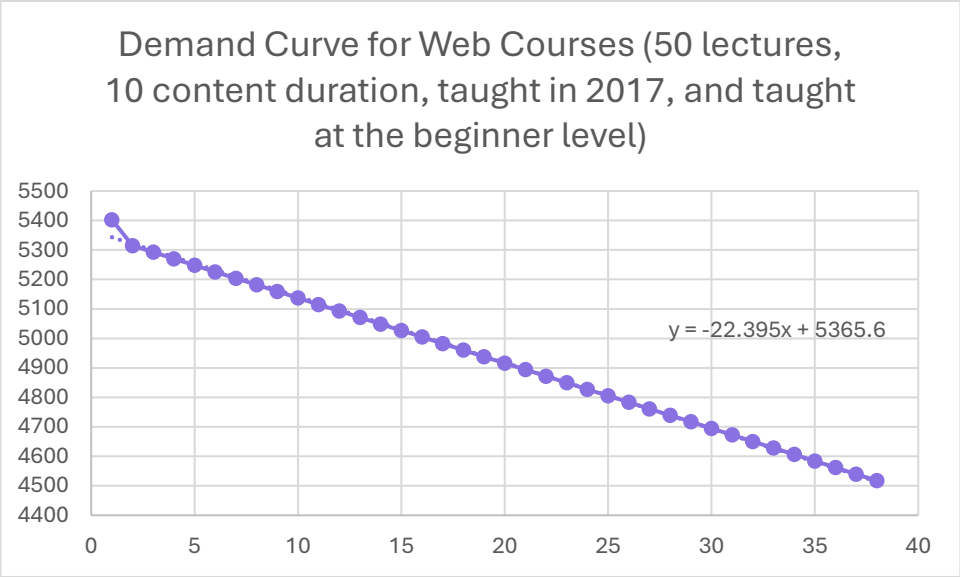
Regression Statistics	
Multiple R	0.479002631
R Square	0.229443521
Adjusted R Square	0.221673203
Standard Error	1.177627284
Observations	1203

	df	*SS*	*MS*	*F*	*Significance F*
Regression	12	491.3986979	40.94989149	29.52820431	1.85652E-59
Residual	1190	1650.299164	1.38680602		
Total	1202	2141.697861			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	9.752	0.619	15.767	0.000	8.538	10.966	8.538	10.966
log_price	-0.154	0.012	-12.443	0.000	-0.179	-0.130	-0.179	-0.130
num_lectures	0.002	0.001	1.537	0.124	0.000	0.004	0.000	0.004
content_duration	0.022	0.009	2.349	0.019	0.004	0.040	0.004	0.040
2012	-0.556	0.593	-0.938	0.349	-1.718	0.607	-1.718	0.607
2013	-0.592	0.552	-1.073	0.284	-1.674	0.491	-1.674	0.491
2014	-1.142	0.540	-2.114	0.035	-2.202	-0.082	-2.202	-0.082
2015	-1.378	0.532	-2.590	0.010	-2.422	-0.334	-2.422	-0.334
2016	-1.752	0.531	-3.299	0.001	-2.794	-0.710	-2.794	-0.710
2017	-2.231	0.534	-4.176	0.000	-3.280	-1.183	-3.280	-1.183
All Level (1)	0.021	0.321	0.066	0.947	-0.608	0.651	-0.608	0.651
Beginner Level(2)	0.152	0.332	0.457	0.647	-0.499	0.802	-0.499	0.802
Expert Level (4)	-0.003	0.319	-0.010	0.992	-0.629	0.623	-0.629	0.623

Question 11

Q. Using your estimates from Q5, draw a linear demand curve for a course with the following characteristics: 50 lectures, 10 content duration, taught in 2017, and taught at the beginner level. Describe what the demand curve shows. *[Hint: remember the regression equation you estimated can be used to predict values of the dependent variable. To draw a linear demand curve, you could predict the quantity for two (or more) values of price and connect the dots.]*



A. From the table Price vs Number of subscribers(on the next page) and demand curve, we can clearly see the downward slope of demand as price increases. This reflects the negative relationship between price and number of subscribers which is consistent with the law of demand. As price increases, the number of subscribers decreases. The slope of the curve gives us insight on the sensitivity to price change. From the regression, coefficient for price was small (-4.43) which suggests low price sensitivity. From this we can understand that increasing prices slightly may not lead to sharp drop in subscribers which is useful for pricing setting strategies. However, for large changes, there will be a more noticeable decrease in subscribers.

Price	Number of Subscribers (Y)
0	5403
20	5314
25	5292
30	5270
35	5248
40	5226
45	5204
50	5182
55	5159
60	5137
65	5115
70	5093
75	5071
80	5049
85	5027
90	5005
95	4982
100	4960
105	4938
110	4916
115	4894
120	4872
125	4850
130	4828
135	4805
140	4783
145	4761
150	4739
155	4717
160	4695
165	4673
170	4651
175	4628
180	4606
185	4584
190	4562
195	4540
200	4518