There are big subgroups in our data men shoes and women's shoes. they are completely different and bundling them together when making predictions is going to yield deceiving results.

Not only feet differ by gender but also there are different shoe types are models.

Our problem is related to inventory management.

Therefore, we should divide our inventory in some way and then count the frequencies. that frequencies will give us a better idea of the data.

A good way to do that is to divide the data by shoe size

we have three Dimensions shoe size, country and gender.

Segment the data by:
1. Shoe size
2. Country
3. Gender

A possible solution is to create two table one for men's shoes and one for women's shoes and then proceed normally

We want to estimate the number of shoes that are likely to be and 95% confidence interval will give us such information.

We will take the last 12 months of sales and make a prediction. let's do this only for men shoes as the problem is identical for both genders.

Frequency distribution tables

Problem: What is the number of shoes that are likely to be sold, based on historical data?

Men sizes

	Country				
US	Canada	United States	United Kingdom	Germany	Total
6	15	54	6	30	**105**
6.5	15	45	12	18	**90**
7	24	39	21	30	**114**
7.5	45	66	12	48	**171**
8	51	141	45	117	**354**
8.5	192	225	87	174	**678**
9	324	492	183	348	**1347**
9.5	375	741	225	549	**1890**
10	237	543	156	411	**1347**
10.5	243	462	150	453	**1308**
11	114	213	69	156	**552**
11.5	75	156	39	129	**399**
12	51	87	24	78	**240**
13	12	39	3	33	**87**
14	21	60	15	30	**126**
15	27	24	12	48	**111**
16	0	0	0	0	**0**

Women sizes

	Country				
US	Canada	United States	United Kingdom	Germany	Total
4	0	0	0	0	**0**
4.5	6	21	15	9	**51**
5	6	9	9	12	**36**
5.5	6	42	6	9	**63**
6	21	33	12	15	**81**
6.5	51	93	24	84	**252**
7	93	147	27	156	**423**
7.5	153	318	87	222	**780**
8	192	618	168	324	**1302**
8.5	171	399	129	339	**1038**
9	213	384	93	264	**954**
9.5	84	189	57	126	**456**
10	48	75	21	87	**231**
10.5	36	87	18	57	**198**
11.5	12	30	3	15	**60**
Total	**1092**	**2445**	**669**	**1719**	**5925**

Game plan: Find the 95% confidence interval using
1. Last 12 months of sales
2. Only for men shoes (as the problem is identical)
3. Only for the USA (as the problem is identical)

actually, we have to calculate 17 confidence intervals for each size. let's get on to it, first we need to calculate the mains.

First, we need to calculate the means

Second, population variance is unknown

One population, population variance unknown -> t-statistic

We have a sample of 12 observations. therefore, we are looking for the t statistics for a 95% confidence interval with 11 degrees of freedom, it is 2.20

	2016
n	12

$$t_{11,0.025} = 2.20$$

Finally, calculate the confidence intervals

$$\bar{x} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}}$$

Frequency distribution tables

By size and month

Problem: What is the number of shoes that are likely to be sold, based on historical data?

Men shoes sales

	United States, 2016												Mean	Standard error	ME	95% CI		Number of pairs
US	1	2	3	4	5	6	7	8	9	10	11	12	2016	2016	2016	2016		
6	4	1	3	1	3	3	3	4	3	7	3	0	2.92	0.51	1.12	1.80	4.04	4
6.5	3	2	0	1	0	0	1	7	2	1	2	1	1.67	0.56	1.21	0.46	2.88	3
7	0	0	1	0	6	4	4	2	3	0	0	0	1.67	0.61	1.32	0.34	2.99	3
7.5	3	2	3	1	7	0	7	3	4	6	1	1	3.17	0.69	1.51	1.65	4.68	5
8	7	9	7	3	12	2	9	4	7	5	2	6	6.08	0.88	1.92	4.16	8.01	8
8.5	12	12	8	8	15	9	17	17	6	9	10	6	10.75	1.12	2.45	8.30	13.20	13
9	17	13	13	11	21	22	25	30	26	25	13	10	18.83	1.97	4.29	14.54	23.12	23
9.5	19	25	27	24	26	33	25	47	31	44	37	26	30.33	2.45	5.33	25.00	35.67	36
10	17	26	26	19	16	31	25	24	23	31	15	20	22.75	1.57	3.42	19.33	26.17	26
10.5	13	16	22	14	28	19	18	15	19	21	16	10	17.58	1.37	2.98	14.60	20.56	21
11	5	16	13	10	10	11	15	8	9	7	6	7	9.75	1.01	2.20	7.55	11.95	12
11.5	4	3	6	3	3	5	6	4	5	12	13	5	5.75	0.96	2.10	3.65	7.85	8
12	3	0	0	4	4	4	3	12	4	9	2	1	3.83	1.01	2.21	1.62	6.04	6
13	1	1	2	0	3	2	1	0	0	4	3	2	1.58	0.38	0.82	0.76	2.41	2
14	2	6	3	3	5	3	2	1	0	1	2	1	2.42	0.50	1.09	1.33	3.50	4
15	0	0	0	1	1	0	4	0	0	0	0	2	0.67	0.36	0.77	-0.11	1.44	1
16	0	0	0	0	0	0	0	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00	0
Total	110	132	134	103	160	148	165	178	142	182	125	98						

	2016
n	12
t_{95}	2.18

The result of obtained can be interpreted as follows.

In 95% of the cases, the true population mean of the sales for each shoe size will fall into the respective interval

The upper bound of the CI shows us the maximum number of pairs needed.
(we have rounded them mathematically, you can instead round them all up if you see fit

mostly we should prepare ourselves with size 9.5 and size 16 won't yield in any sales.

sorry for those of you who are size 16 feet. I know it is hard enough to find shoes already but while this company would not be selling any.

- **we can see two tables representing the sales of women shoes into German shops**

now an assumption that we have to make is that the same people don't buy pairs of shoes form different shops

Assumption: same people don't buy shoes from different shops in the same year

logically it makes sense that in the same year the same people don't go around different shops of the same brand buy shoes. Even if this happens it is an exception and not the norm.

The two samples are independent

The two samples are independent, population variance unknown, but assumed equal

Therefore, we can say that the two samples are independent, once again we don't know the population variance but given that this is the same market in the same country, we can assume it has equal.

Frequency distribution tables
By size and month

Problem: By how much one shop outperforms the other in terms of sales?

Women shoe sales

	Germany, GER1												Germany, GER2												Mean		Sample variance		Pooled variance	Margin of error	95% CI	
US	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	GER1	GER2	GER1	GER2				
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4.5	0	0	0	0	1	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0.42	0.08	0.81	0.08	0.45	0.57	-0.23	0.90
5	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0.17	0.17	0.33	0.33	0.33	0.49	-0.49	0.49
5.5	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	2	0	1	0.08	0.33	0.08	0.42	0.25	0.43	-0.68	0.18
6	0	2	0	0	0	0	0	0	0	0	0	0	0	1	3	1	2	0	0	0	0	0	0	0	0.17	0.58	0.33	0.99	0.66	0.69	-1.11	0.27
6.5	3	3	1	2	1	0	2	0	2	1	3	4	2	0	2	1	1	2	0	1	2	1	3	0	1.83	1.25	1.61	0.93	1.27	0.95	-0.37	1.54
7	0	3	3	4	1	0	1	0	2	0	0	1	0	0	0	4	1	3	1	1	1	3	1	4	1.25	1.58	2.02	2.27	2.14	1.24	-1.57	0.91
7.5	1	2	4	1	2	6	4	3	5	8	2	1	2	1	1	3	2	7	9	8	14	8	6	3	3.25	5.33	4.93	16.06	10.50	2.74	-4.83	0.66
8	6	10	3	9	1	3	6	8	3	12	3	9	13	6	5	13	5	3	11	6	6	9	8	3	6.08	7.33	12.27	12.24	12.25	2.96	-4.21	1.71
8.5	10	10	10	7	14	4	7	7	4	8	7	9	8	5	10	4	5	5	9	7	3	7	9	8	8.08	6.67	7.72	4.97	6.34	2.13	-0.72	3.55
9	1	3	8	6	3	1	4	4	0	2	4	2	5	2	2	9	3	1	1	7	2	1	4	2	3.17	3.25	5.06	6.57	5.81	2.04	-2.13	1.96
9.5	4	1	2	1	2	2	2	4	5	2	3	2	0	1	1	0	1	2	2	1	7	2	4	2	2.50	1.92	1.55	3.72	2.63	1.37	-0.79	1.96
10	0	1	1	1	1	1	3	1	0	0	0	1	0	1	1	0	0	0	2	3	0	2	0	0	0.83	0.75	0.70	1.11	0.91	0.81	-0.72	0.89
10.5	1	0	0	0	2	2	4	1	0	3	1	1	0	2	0	0	0	1	0	0	0	0	2	1	1.25	0.50	1.66	0.64	1.15	0.91	-0.16	1.66
11.5	0	0	0	2	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	1	0	0	0	0	0.17	0.50	0.33	2.09	1.21	0.93	-1.27	0.60
Total	26	35	32	33	28	22	35	28	21	36	25	30	30	19	30	35	20	24	35	38	35	36	37	24								

	GER1	GER2
n	12	12

$t_{95\%,22}$ 2.07

All confidence intervals start in the negatives and finish in the positives

We cannot conclude one shop sells more shoes than the other for any size

For some sizes, GER1 is likely to sell more, while for others - vice versa (you will check that for yourself for homework)

Insight: these two shops are so balanced in terms of sales, they may be bundled together

On average, they will move together. They are predicted to remain identical