PROJECT 3

Kankanamge Harsha

A)Latin Square Design (LSD)

A study is conducted to determine if there are differences in the average prices among four major supermarkets in a given city identified as (A, B, C, and D). In order to eliminate any effects due to different items and different brands, four regularly bought items such as cereal, peanut butter, cookies and detergent and four different brands of each item available at each supermarket were selected. Prices (in dollars) are recorded using a Latin Square Design in **641PROJ3A.sav**.

a) Identify the response variable, treatments and blocks. Write the model for the LSD experiment. What is the value of 'r'?

Response variable: Prices

Treatment: Four major supermarkets (A, B, C, D)

Blocks: **Items** (Cereal, peanut butter, cookies and detergent) and **four brands**

α_i Effect of i^{th} row
$ß_j$ Effect of j^{th} column
τ_k Effect of k^{th} treatment
μ ground mean
$\mathcal{E}_{ijk}$ Random error

$Y_{ijk} = \mu_{ijk} + \mathcal{E}_{ijk}$

i = 1,2,3,4
j = 1,2,3,4
k = 1,2,3,4

$$y_{ijk} = \mu + \alpha_i + ß_j + \tau_k + \mathcal{E}_{ijk}$$

Where μ ground mean
Row effect α_i are such that $\sum_{i=1}^{4} \boldsymbol{\alpha i} = 0$
Column effect $ß_j$ such that $\sum_{j=1}^{4} ß\mathbf{j} = 0$
Treatment effect τ_k such that $\sum_{k=1}^{4} \tau k = 0$

Value of r is 4

b) Test if 'prices' are normally distributed among the four supermarkets.

H0: 'Price' is normally distributed among all four supermarket
H1: 'Price' is not normally distributed among all four supermarket

Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	4 supermarkets	Statistic	df	Sig.	Statistic	df	Sig.
price in dollars	A	.243	4	.	.923	4	.553
	B	.247	4	.	.912	4	.494
	C	.238	4	.	.931	4	.601
	D	.245	4	.	.921	4	.543

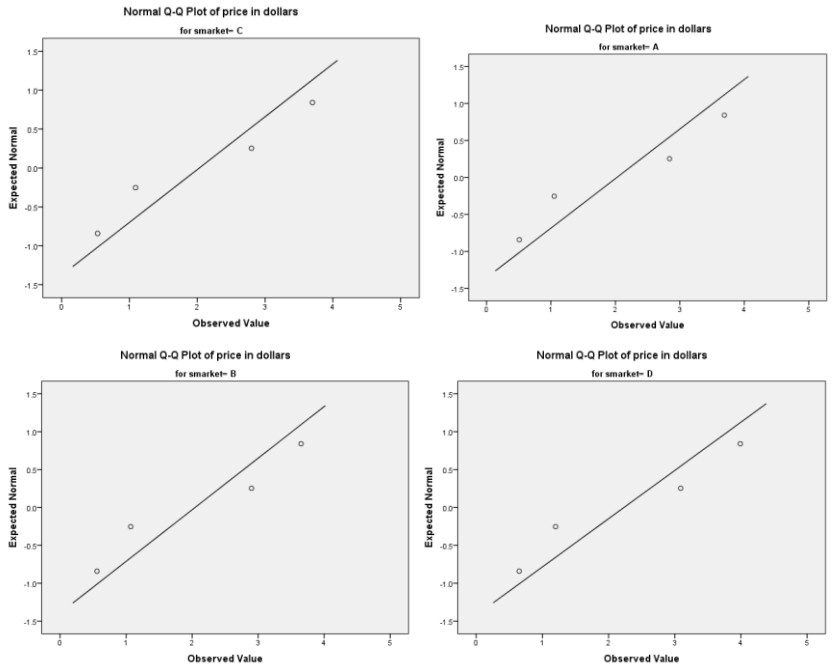
a. Lilliefors Significance Correction

Prices among:

Supermarket A: Shapiro-Wilk test statics 0.923 and p-value 0.553 >0.01
Supermarket B: Shapiro-Wilk test statics 0.912 and p-value 0.494 >0.01
Supermarket C: Shapiro-Wilk test statics 0.931 and p-value 0.601 >0.01
Supermarket D: Shapiro-Wilk test statics 0.921 and p-value 0.543 >0.01

Do not reject the H0

At 0.01 level of significant price is normally distributed among all four supermarket



All the data points are fitting onto the line. Which mean ‘prices’ data are randomly distributed among all four supermarket

Descriptives

	4 supermarkets			Statistic	Std. Error
price in dollars	A	Mean		2.0225	.74623
		99% Confidence Interval for Mean	Lower Bound	-2.3362	
			Upper Bound	6.3812	
		5% Trimmed Mean		2.0139	
		Median		1.9450	
		Variance		2.227	
		Std. Deviation		1.49246	
		Minimum		.51	
		Maximum		3.69	
		Range		3.18	
		Interquartile Range		2.83	
		Skewness		.161	1.014
		Kurtosis		-3.889	2.619
	B	Mean		2.0450	.73388
		99% Confidence Interval for Mean	Lower Bound	-2.2415	
			Upper Bound	6.3315	
		5% Trimmed Mean		2.0383	
		Median		1.9850	
		Variance		2.154	
		Std. Deviation		1.46775	
		Minimum		.56	
		Maximum		3.65	
		Range		3.09	
		Interquartile Range		2.78	
		Skewness		.118	1.014
		Kurtosis		-4.208	2.619
	C	Mean		2.0300	.73685
		99% Confidence Interval for Mean	Lower Bound	-2.2739	
			Upper Bound	6.3339	
		5% Trimmed Mean		2.0206	
		Median		1.9450	
		Variance		2.172	
		Std. Deviation		1.47370	
		Minimum		.53	
		Maximum		3.70	
		Range		3.17	
		Interquartile Range		2.81	
		Skewness		.189	1.014
		Kurtosis		-3.625	2.619
	D	Mean		2.2325	.78499
		99% Confidence Interval for Mean	Lower Bound	-2.3525	
			Upper Bound	6.8175	
		5% Trimmed Mean		2.2228	
		Median		2.1450	
		Variance		2.465	
		Std. Deviation		1.56998	
		Minimum		.65	
		Maximum		3.99	
		Range		3.34	
		Interquartile Range		2.98	
		Skewness		.171	1.014
		Kurtosis		-3.904	2.619

c)Test if 'prices' have homogeneous variance (Use Cochran's Test) for each of the four supermarkets.

H0: $\sigma^2 1 = \sigma^2 2 = \sigma^2 3 = \sigma^2 4$
H1: At least one of the σ^2_i of price is different among the four supermarket.

$S^2 1$ (A)= 2.227
$S^2 2$ (B)= 2.154
$S^2 3$ (C)= 2.172
$S^2 4$ (D)= 2.465

Total of S^2_i = 9.018

$$G = \frac{\text{largest } S_i^2}{\sum_{i=1}^{k} S_i^2},$$

G=2.465/9.018 = 0.27334 g (0.01 k=4) = 0.7814

G (0.27334) <g (0.7814) Do not reject H0

All variances of 'prices' are same among the all four supermarket.

d)Using an LSD, test for differences in prices among the 4 supermarkets.

H0: $\tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$
H1: At least one of treatment (Supermarket) effect (τ_i) is not equal to zero.

Tests of Between-Subjects Effects

Dependent Variable: price in dollars

Source	Type II Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	27.163[a]	9	3.018	1425.900	.000
Intercept	69.389	1	69.389	32782.157	.000
items	27.028	3	9.009	4256.409	.000
smarket	.121	3	.040	19.063	.002
brand	.014	3	.005	2.228	.186
Error	.013	6	.002		
Total	96.565	16			
Corrected Total	27.176	15			

a. R Squared = 1.000 (Adjusted R Squared = .999)

Supermarket:

F statistic = 19.063 p-value = 0.002

α (0.01)> p-value (0.002)

Reject H0

At least one of treatment (Supermarket) effect (τ_i) is not equal to zero. Which means least one mean value of 'price' among four supermarkets is different.

e)Conduct appropriate Tukey and Bonferroni tests for multiple comparisons.

Multiple Comparisons

Dependent Variable: price in dollars

	(I) 4 supermarkets	(J) 4 supermarkets	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	A	B	-.0225	.03253	.897	-.1351	.0901
		C	-.0075	.03253	.995	-.1201	.1051
		D	-.2100*	.03253	.003	-.3226	-.0974
	B	A	.0225	.03253	.897	-.0901	.1351
		C	.0150	.03253	.965	-.0976	.1276
		D	-.1875*	.03253	.005	-.3001	-.0749
	C	A	.0075	.03253	.995	-.1051	.1201
		B	-.0150	.03253	.965	-.1276	.0976
		D	-.2025*	.03253	.003	-.3151	-.0899
	D	A	.2100*	.03253	.003	.0974	.3226
		B	.1875*	.03253	.005	.0749	.3001
		C	.2025*	.03253	.003	.0899	.3151
Bonferroni	A	B	-.0225	.03253	1.000	-.1482	.1032
		C	-.0075	.03253	1.000	-.1332	.1182
		D	-.2100*	.03253	.004	-.3357	-.0843
	B	A	.0225	.03253	1.000	-.1032	.1482
		C	.0150	.03253	1.000	-.1107	.1407
		D	-.1875*	.03253	.007	-.3132	-.0618
	C	A	.0075	.03253	1.000	-.1182	.1332
		B	-.0150	.03253	1.000	-.1407	.1107
		D	-.2025*	.03253	.005	-.3282	-.0768
	D	A	.2100*	.03253	.004	.0843	.3357
		B	.1875*	.03253	.007	.0618	.3132
		C	.2025*	.03253	.005	.0768	.3282

Based on observed means.
The error term is Mean Square(Error) = .002.

*. The mean difference is significant at the .05 level.

H0: $\mu_i = \mu_j$ i, j=1,2,3,4
H1: $\mu_i \neq \mu_j$ $i \neq j$

α=0.01

Tukey HSD: Means differences of "prices "between supermarket D and A, B, C are significant. Which means:

Supermarket: D – A price differences $ 0.21 p-value =0.003<0.01 Reject H0. Significant.
Supermarket: D – B price differences $ 0.1875 p-value =0.005<0.01 Reject H0. Significant.
Supermarket: D – C price differences $ 0.2025 p-value =0.003<0.01 Reject H0. Significant.

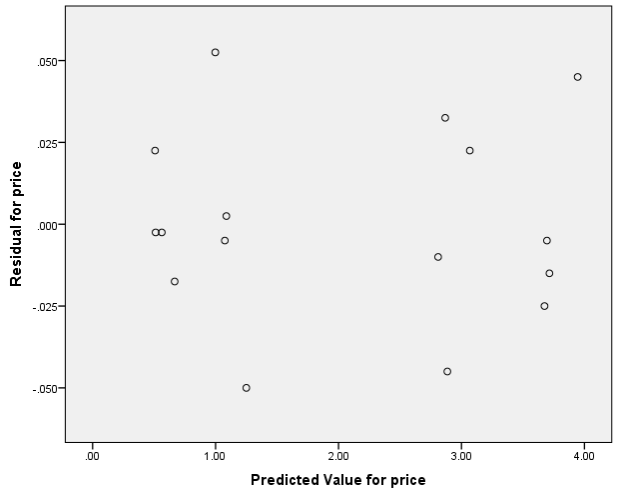
Means differences of “prices “in between supermarkets A, B and C are not significant. P-value (closer to 0.9) > 0.01 (Do not reject H0) and differences are very low.

Bonferroni: Means differences of “prices “between supermarket D and A, B, C are significant. Which means:

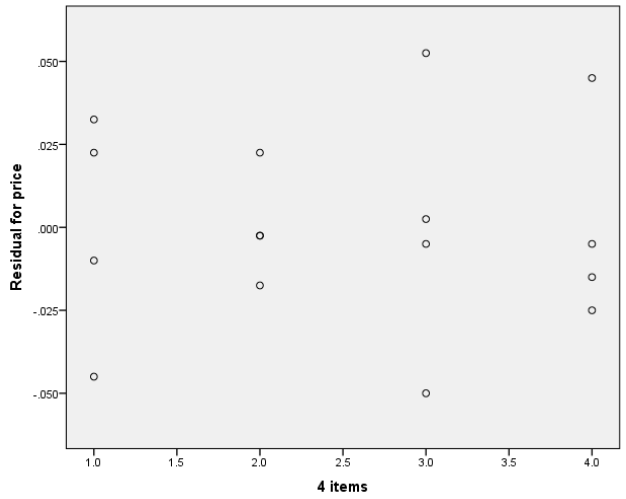
Supermarket: D – A price differences $ 0.21 p-value =0.004<0.01 Reject H0. Significant.
Supermarket: D – B price differences $ 0.1875 p-value =0.007<0.01 Reject H0. Significant.
Supermarket: D – C price differences $ 0.2025 p-value =0.005<0.01 Reject H0. Significant.

Means differences of “prices “in between supermarkets A, B and C are not significant. P-value =1> 0.01 (Do not reject H0) and differences are very low.

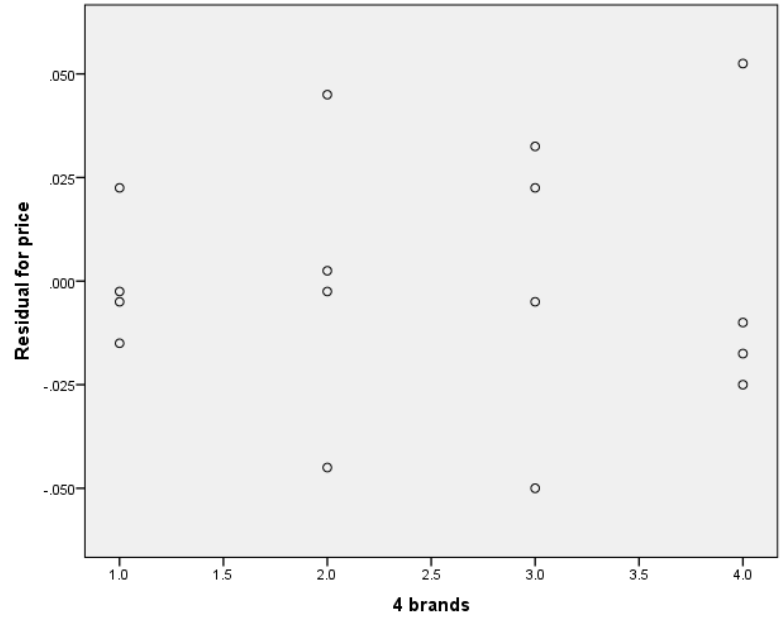
f)Comment on the four residual plots.



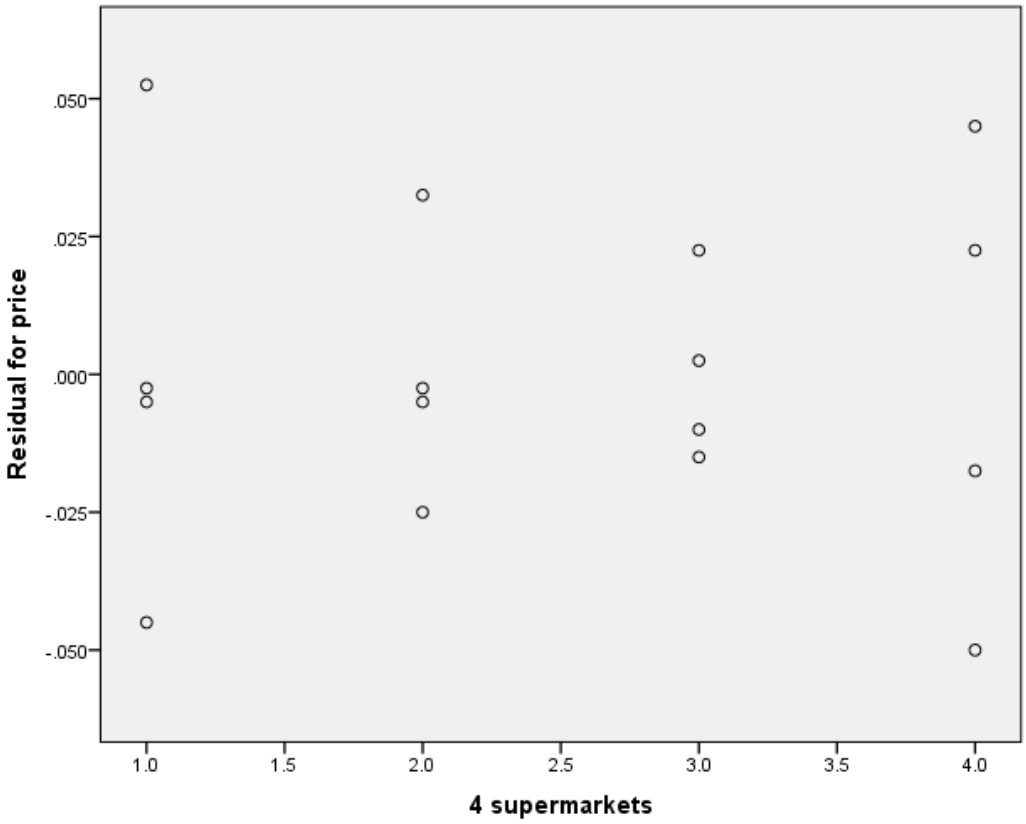
The residual plot (predicted value for price vs residual for price) show that most of the data points are within the band -0.05 to +0.05. As well as residuals don’t shows pattern. Which means plot of residuals verses predicted price indicate error variance of residuals may be the same for the different predicted prices. Therefore, variances of residuals are homogeneous with respect to predicted price.



The residual plot (items vs residual for price) show that most of the points are within the band -0.05 to +0.05. But residuals are showing a pattern. Error variance of residuals with respect to item 2 is very low compare with other items. And also error variance of residuals with respect to item 3 is very high compare with other items. Which means plot of residuals verses items indicate error variance of residuals may not be the same for the different items. Therefore, variances of residuals are not homogeneous with respect to items.

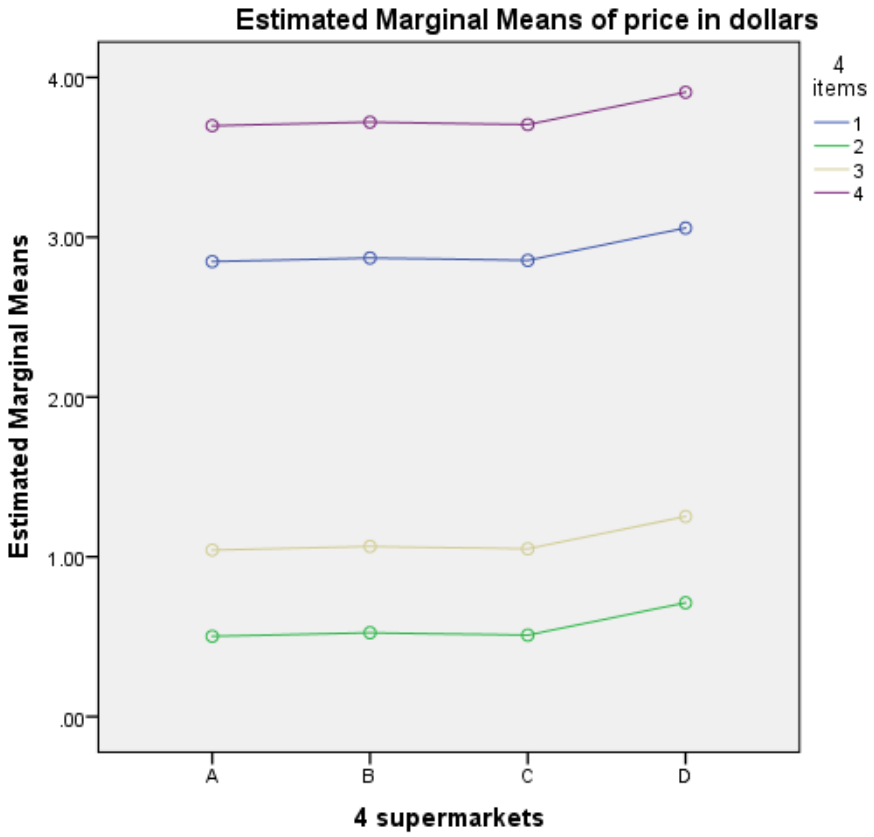


The residual plot (brands vs residual for price) show that most of the points are within the band -0.05 to +0.05. But residuals are showing a pattern. Error variance of residuals for brand 1 is very low compare with other brands. Which means plot of residuals verses brands indicate error variance of residuals may not be the same for the different brands. Therefore, variances of residuals are not homogeneous with respect to brands.



The residual plot (Supermarket vs residual for price) show that most of the data points are within the band -0.05 to +0.05. But residuals are showing a pattern. Error variance of residuals for supermarket C (3) is very low compare with other supermarket. Error variance of residuals for supermarket A (1) and D (4) is very high compare with other supermarkets. Which means plot of residuals verses supermarkets indicate error variance of residuals may not be the same for the different supermarkets. Therefore, variances of residuals are not homogeneous with respect to supermarket.

g)Comment on the plots of mean price by supermarket separately for each block of the two blocking variables.



There are not price interactions of four items among four supermarkets.
Item four has highest price on mean among all supermarket compare with other items and also supermarket D has higher price for an item four compare with other supermarkets.
We can simply say that item four is expensive compared with other items.

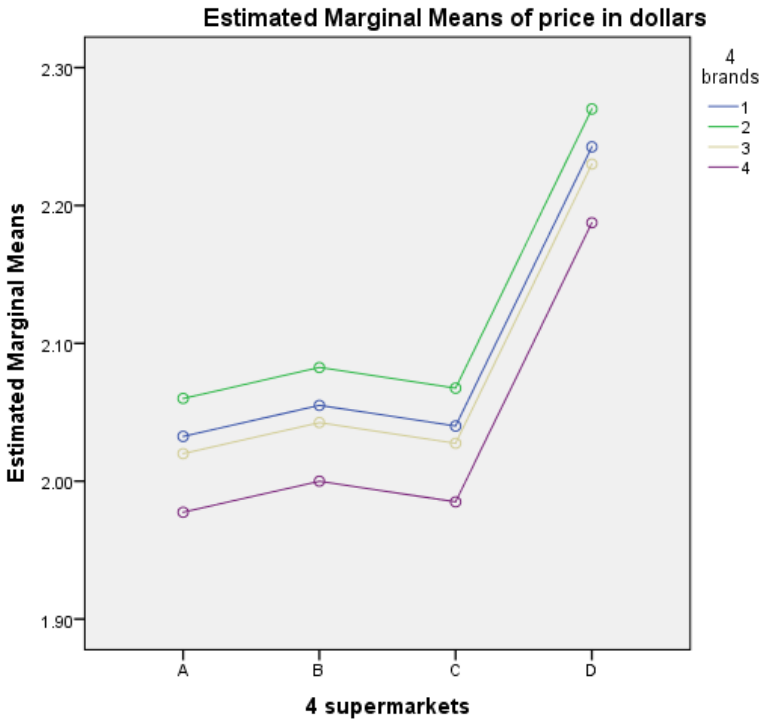
Supermarket D has highest price on mean for all four items compared with other supermarket.
Supermarkets A and C have lowest mean price for all four items compared with other supermarkets.
We can simply say that for same items supermarket D has higher price and supermarkets A and C have lower price on mean.

Item two is the cheapest item in all supermarket.

Supermarket A, B, C have very closer price on mean for same items. But supermarket D has higher price on mean for same items.

Item 1 is the second highest expensive item in all four supermarkets.

All the items are expensive in supermarket D compared with other three supermarkets.



There are not price interactions of four brands among four supermarkets.
Brand two has highest price on mean among all supermarket compare with other brands and also supermarket D has higher price for brand 2 compare with other supermarkets.
We can simply say that brand 2 is expensive compared with other brands and also it is very expensive in supermarket D compared with other three supermarkets.

Supermarket D has highest price on mean for all four brands compared with other supermarket.
Supermarket A and C have lowest prices on means for all four brands compared with other supermarkets.
We can simply say that for same brands supermarket D has higher prices and supermarket A and C have lower prices on mean.

Brand four is the cheapest brand in all supermarkets.

Supermarket A, C have very closer price on mean for same items and supermarket B has little higher price on mean than A and C. But supermarket D has higher prices on mean for same brands.

Brand 1 is the second highest expensive brand in all four supermarkets.

Supermarket D has higher price on mean for all brands and items among all four supermarkets. Which means regardless of brands or items stuffs in supermarket D is expensive compared with other supermarkets. So supermarket D has significant difference on prices compared with other three supermarkets.

But Supermarkets A, B, and C have very closer prices on mean for all brands and items. Which means there are not significant differences on mean prices in supermarkets A, B and C.

B)Among nine persons interviewed in a poll, three are Easterners, three are Southerners, and three are Westerners. By profession, three of them are teachers, three are lawyers, and three are doctors, and no two of the same profession come from the same part of the United States. Also, three are Democrats, three are Republicans, and three are Independents, and no two of the same political affiliation are of the same profession or come from the same part of the United States. If one of the teachers is an Easterner and an Independent, another teacher is a Southerner and a Republican, and one of the lawyers is a Southerner and a Democrat, what is the political affiliation of the doctor who is a Westerner? **(Hint: Construct a Latin Square with r = 3, remembering that every entry occurs exactly once in each row and once in each column).**

	Teachers	Lawyers	Doctors
Easterners	Independent	Republican	Democrat
Southerners	Republican	Democrat	Independent
Westerners	Democrat	Independent	Republican

Doctor who is a Westerner is Republican