California State University – East Bay

MS in Business Analytics
BAN 602 Case 4 Solution
Group 5

1. Aastha Tandon atandon@horizon.csueasybay.edu
2. Sukirti Bhat sbhat2@horizon.csueasybay.edu
3. Venu Madhava Reddy Kotla vkotla@horizon.csueastbay.edu
4. Shaik Ahamed sahamed@horizon.csueastbay.edu

Professor:
Somak Paul

1. Sales professionals in SFO conducted a survey of its membership to study the relationship b/w years of experience & salary of employees in inside and outside sales positions. Below are the descriptive statistics for summarizing the compensation data by experience and sales positions.

Table 1.1 Salary based on Experience									
Experience	**Mean**	**Median**	**Variance**	**Standard Deviation**	**Q1**	**Q3**	**Maximum**	**Minimum**	**Range**
Low	59820	60410	36060691	6005.055	55447	63813	71345	48621	22724
Medium	68618	69616	185541915	13621.377	54646	81827	88730	51246	37484
High	66339	65441	94080544	9699.513	57771	76975	79081	51027	28054

Table 1.1 represents the Salary structure based on three levels of years of experience: low (1-10 years), medium (11-20 years), and high (21 or more years).
For medium experience group, mean and median salary is the highest followed by high experience and then low experience group. As we can see from the table 1.1, the medium group has the highest variance as compared to high and low years of experience which means data is highly spread for medium exp. group. Additionally, the maximum salary is also received by the medium years of experience employee in comparison to others. The IQR of the medium group is also the largest followed by high and low group.

Table 1.2 Salary based on Position									
Position	**Mean**	**Median**	**Variance**	**Standard Deviation**	**Q1**	**Q3**	**Maximum**	**Minimum**	**Range**
Inside	56021	56210	12886852	3589.826	53191	58202	64562	48621	15941
Outside	73830	76316	62773302	7922.96	67008	78972	88730	60259	28471

Table 1.2 represents salary structure for individuals employed in inside and outside sales positions irrespective of their years of experience. From table 1.2, we can see that the mean and median salary of outside positioned employees are more than the inside employees. The variance and Standard deviation and IQR are also higher for the outside than inside.
This is **not a controlled experiment** since no factors are controlled and it does not influence the variables of interest. Infact, this is an observational study where data are usually obtained through sample surveys.

2. As per the data given, mean annual salary for all salesperson is = **64925.47** ($\bar{x}$)
Variance = **117476832** (s^2)
Standard deviation = **10838.67** (s)
We want to develop a 95% confidence interval estimate for the mean salary of salespersons. 95% confidence means (1-α) = 0.95. α = 0.05. $\frac{\alpha}{2}$ = 0.025. Degree of freedom (DOF) = (n-1). Here sample size n = 120. Hence **DOF = 119.** From t table we get t = 1.984 for 100 DOF. In T table there is no value for DOF 119. Hence taking the nearest value.
<u>Formula for interval estimate when population variance is not known:</u>
$\bar{x} \pm t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$ = 64925.47 $\pm$ 1.984 ($\frac{10838.67}{\sqrt{120}}$) = 64925.47 $\pm$ 1963.03
Intervals: 62961.97 - 66888.03
Note: Exact interval values will be calculated through R (62966.30 - 66884.65)

3. We are performing analysis of variance to test for any significant differences in annual salary due to years of experience, ignoring the effect of position.
Following are the assumptions of Analysis of variance:
1. For each population, the response (dependent) variable is normally distributed.
2. The variance of the response variable, denoted σ2, is the same for all the populations.
3. The observations must be independent.

Table 3.1

Experience	Low	Medium	High
Mean	59819.62 ($\bar{x}_1$)	68618.12 ($\bar{x}_2$)	66338.68 ($\bar{x}_3$)
Variance	36060691 (${s_1}^2$)	185541915(${s_2}^2$)	94080544 (${s_3}^2$)

Table 3.1 shows sample mean and variance among different years of experience group.

Factor: Variable selected for investigation in an experiment. Here factor is **Experience**.
Treatments: It is levels of factor. Here treatments are: **Low, Medium, High**
Experimental units: This is object of interest in experiment - **Sales professionals**
Response variable: **Annual salary**
Let,
$\mu 1$ = Population Mean of low years of experience group
$\mu 2$ = Population Mean of medium years of experience group
$\mu 3$ = Population Mean of high years of experience group

H_0 (Null Hypothesis): $\mu 1 = \mu 2 = \mu 3$
H_a (Alternative Hypothesis): Not all population means are equal
Level of significance provided is $\boldsymbol{\alpha}$ **= 0.05**

Now let us compute the value of test statistics Where:
n_j = number of observations for treatment j = 40
$\bar{x}_j$= sample mean for treatment j (Refer table 3.1)
${s_j}^2$ = Sample variance for treatment j. (Refer table 3.1)
$\bar{\bar{x}}$= Overall sample mean= $\frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3}{3} = \frac{66338.68 + 59819.62 + 68618.12}{3}$ = **64925.47**
Mean square due to treatments: The estimate of σ2 based on the variation of the sample means is called the mean square due to treatments and is denoted by MSTR. Numerator is called the sum of squares due to treatments (SSTR). Denominator is the degrees of freedom associated with SSTR which is k-1= 3-1 = 2
SSTR: $\sum_{j=1}^{k} n_j(\bar{x}_j - \bar{\bar{x}})^2 = 40(59819.62 - 64925.47)^2 + 40(68618.12 - 64925.47)^2 + 40(66338.68 - 64925.47)^2$ = **1668101230**
MSRT: $\frac{SSTR}{(k-1)} = \frac{1668101230}{2}$ = **834050615**

Mean square due to error: The estimate of σ2 based on the variation of the sample observations within each sample is called the mean square error and is denoted by MSE. Numerator is called the sum of squares due to error (SSE). Denominator is the degrees of freedom associated with SSE which is $n_T - k$ = (120 – 3) = **117**
SSE=$\sum_{j=1}^{k}(n_j - 1){s_j}^2 = 39 * 36060691 + 39 * 185541915 + 39 * 94080544$ = **12311642850**

MSE = $\frac{SSE}{n_T - k} = \frac{12311642850}{117}$ = **105227716.7**

F = $\frac{MSTR}{MSE}$ = **7.92**

Critical Value Approach:

We will reject H_0 if F > $F_{0.05}$

Per F table, $F_{0.05}$ is less than 3.09. With F = 7.92, F > $F_{0.05}$. We will reject H_0

Table 3.2 ANOVA Table					
Source of Variation	**Sum of Square**	**Degree of Freedom**	**Mean Square**	**F**	**p-value**
Treatment	1668101230	2	834050615	7.92	less than 0.01
Error	12311642850	117	105227716.7		
Total	13979744080	119			

4. We want to test differences in annual salary due to position ignoring effect of experience, using analysis of variance test. We have calculated mean and variance of 60 inside and 60 outside positions in Table 4.1

Table 4.1

	Inside	**Outside**
Mean	56020.52 ($\bar{x}_1$)	73830.43($\bar{x}_2$)
Variance	12886852(${s_1}^2$)	62773302(${s_2}^2$)

Factor: Positions
Treatments: Inside and Outside
Experimental units: Sales professionals
Response variable: Annual salary

$H_0 : \mu_1 = \mu_2$ (μ_1 = Mean of inside position and μ_2= Mean of outside position)
$H_a : \mu_1 \neq \mu_2$
Level of significance provided is $\boldsymbol{\alpha}$ **= 0.05**
Now let us compute the value of test statistics where:
n_j = number of observations for treatment j = **60**
$\bar{x}_j$= sample mean for treatment j (Refer table 4.1)
$\bar{\bar{x}}$ = Overall sample mean
${s_j}^2$ = Sample variance for treatment j (Refer table 4.1)
k-1 = Degree of freedom for SSTR = 2-1 = **1**
$n_T - k$ = Degree of freedom for MSE = (120 – 2) = **118**

Mean square due to treatments:

$\bar{\bar{\boldsymbol{x}}} = \frac{\bar{x}_1 + \bar{x}_2}{2} = \bar{\bar{x}} = \frac{56020.52 + 73830.42}{2}$ = **64925.48**

SSTR = $\sum_{j=1}^{k} n_j (\bar{x}_j - \bar{\bar{x}})^2$
= $60(56020.52 - 64925.48)^2 + 60(73830.43 - 64925.48)^2$ = **9515786826**
MSTR = $\frac{SSTR}{(k-1)} = \frac{9515786826}{(2-1)}$ **= 9515786826**

Mean square due to error:

SSE = $\sum_{j=1}^{k}(n_j - 1)s_j^{\,2}$ = 59(12886852) + 59(62773302) = **4463949086**

MSE = $\frac{SSE}{(120-2)}$ = **37830077**

F = $\frac{MSTR}{MSE}$ = **251.54**

Critical Value Approach:

We will reject H_0 if F > $F_{0.05}$

Per F table, $F_{0.05}$ will be less than 3.94. With F = 251.54, F > $F_{0.05}$. We will reject H_0

Table 4.1: ANOVA Table

Source of Variation	Sum of Square	Degree of Freedom	Mean Square	F	p-value
Treatment	9515786826	1	9515786826	251.54	<0.01
Error	4463949086	118	37830077		
Total	13979735912	119			

5. We want to test for any significant differences in annual salary due to position, years of experience, and the interaction of those two, at 0.05 level of significance.

Factor A: Position type (2 Levels) = a = **2**

Factor B: Experience (3 Levels) = b = **3**

Replication: Each experiment condition is repeated 20 times = r = **20**

Total number of observations taken in experiment n_T = abr = **120**

Table 5.1

Treatment combination Total / **Sample mean for observation**		**Factor B: Experience** Low	Medium	High	**Row Total**	**Factor A means**
Factor A: Position	Inside	1100627	1112155	1148449	3361231	56020.52
		55031.35	55607.75	57422.45		
	Outside	1292158	1632570	1505098	4429826	73830.43
		64607.9	81628.5	75254.9		
Column Totals		2392785	2744725	2653547	7791057	**← Overall Total**
Factor B means		59819.625	68618.125	66338.675	64925.475	**← Overall sample mean of all observation**

Notations:

$\boldsymbol{x_{ijk}}$ = observation corresponding to the kth replicate taken from treatment i of Position (Factor A) and treatment j of Experience (factor B). In table 5.1, x_{ijk} is 20 observations corresponding to every Factor A and every Factor B. Count of x_{ijk} is 120 total observations.

Note: Since there are total of 120 observations, we are not including that in report.

$\boldsymbol{\bar{x}_{i.}}$ = Sample mean of observation in treatment i (factor A).

In table 5.1, $\bar{x}_{1.}$ = 56020.52, $\bar{x}_{2.}$ = 73830.43

$\boldsymbol{\bar{x}_{.j}}$ = Sample mean of observation in treatment j (factor B).

In table 5.1, $\bar{x}_{.1}$ = 59819.625, $\bar{x}_{.2}$ = 68618.125, $\bar{x}_{.3}$ = 66338.675

$\boldsymbol{\bar{x}_{ij}}$ = sample mean of observation which corresponds to i and j treatments respectively.

In table 5.1, $\bar{x}_{11}$ = 55031.35, $\bar{x}_{12}$ = 55607.75, $\bar{x}_{13}$ = 57422.45, $\bar{x}_{21}$ = 64607.9, $\bar{x}_{22}$ = 81628.5, $\bar{x}_{23}$ = 75254.9

$\boldsymbol{\bar{\bar{x}}}$ = Overall sample means of all observations. In Table 5.1, $\bar{\bar{x}}$ = **64925.475**

Step1: Compute total sum of squares SST:

SST = $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(\bar{x}_{ijk} - \bar{\bar{x}})^2$

$= (57718 - 64925.475)^2 + (53938 - 64925.475)^2 \ldots\ldots (67603 - 64925.475)^2$

=**13979742992**

Step2 & 3: Compute sum of squares for factor A and factor B:

SSA = $br\sum_{i=1}^{a}(\bar{x}_{i.} - \bar{\bar{x}})^2 = 60\,[(56020.52 - 64925.475)^2 + (72830.43 - 64925.475)^2]$

= **9515786826**

SSB = $ar\sum_{j=1}^{b}(\bar{x}_{.j} - \bar{\bar{x}})^2$

$= 40\,[(59819.625 - 64925.475)^2 + (68618.125 - 64925.475)^2 + (66338.675 - 64925.475)^2]$

=**1668100099**

Step4: Compute sum of squares for interaction:

SSAB = $r\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{\bar{x}})^2$

$=20[(55031.35 - 56020.52 - 59819.625 + 64925.475)^2 \ldots\ldots (75254.9 - 73830.43 - 66338.675 + 64925.475)^2]$ = **1352066184**

Step5: Compute sum of squares due to error:

SSE = SST – SSA – SSB - SSAB

= 13979742992 – 9515786826 – 1668100099 – 1352066184 = **1443789883**

Step6: Computing mean square:

MSA = $\frac{SSA}{a-1} = \frac{9515786826}{2-1}$ = **9515786826**

MSB = $\frac{SSB}{b-1} = \frac{1668100099}{3-1}$ = **834050049.5**

MSAB = $\frac{SSAB}{(a-1)(b-1)} = \frac{1352066184}{(1)(2)}$ = **676033092**

MSE = $\frac{SSE}{ab(r-1)} = \frac{1443789883}{6(20-1)} = \frac{1443789883}{114}$ = **12664823.54**

Table 5.2 ANOVA table for Two- Factor Factorial Experiment

SOURCE	Sum of Square	Degree of Freedom	Mean Square	F	p-value
Factor A	9515786826	1	9515786826	751.36	very less than 0.01
Factor B	1668100099	2	834050049.5	65.86	very less than 0.01
Interaction	1352066184	2	676033092	53.38	very less than 0.01
Error	1443789883	114	12664823.54		
Total	13979742992	119			

Level of Significance provided α = 0.05.

Critical value approach:

- Position: F = 751.36. $F_{0.05}$ = less than 3.94. Since F ≥ $F_{0.05}$, mean annual salary differs by position
- Experience: F = 65.86, $F_{0.05}$ = less than 3.94, Since F ≥ $F_{0.05}$, mean annual salary differs by experience.
- Interactions: F = 53.38, $F_{0.05}$ = less than 3.94, Since F ≥ $F_{0.05}$, Interaction is significant.