CASELETS OF FINANCIAL ANALYTICS

MINI CASE STUDIES OF AN ADVISORY FIRM IN SINGAPORE

PART I: STATISTICAL ANALYSIS

I. TABLE OF CONTENTS

II. LIST OF FIGURES

III. INTRODUCTION

This manuscript accounts for instances in which I, Dr. Alvin Ang, assisted Ms. Joey Wu in analyzing and developing her team of Financial Advisors.

Currently, Ms. Joey Wu is taking charge of some Independent Financial Advisors (IFA) under her care; in which she is solely responsible for their sales results. I have had the honor of operationalizing her teams' data and drew some insights from our analytics. Basically, I was labeled as the team's consultant cum analyst cum sales trainer.

However, names & figures had to be altered for the sake of confidentiality. In such cases, we simulated data akin to what we observed on the grounds.

It is the hope that our work here will pave the way for greater success as Joey's team grows. In other words, we hope that this manuscript will shorten the understanding of her team's operations as new members join her in future.

IV. A LOOK AT THE DISTRIBUTION OF MONTHLY COMMISSIONS FOR NEW CONSULTANTS

A. SYNOPSIS

It is well known within the Financial Advisory (FA) industry that the initial performance of new consultants is of vital importance. This is because new consultants are starting off on a clean sheet and are unable to receive renewal commissions. In other words, for most new consultants, their customer base is small, and so is their sales commission (compared to senior consultants who have already built up their customer base over the years, and are able to sustain a better income thru renewal commissions).

Without a good starting commission as their salary, new consultants may find it tough to sustain, and drop out of their careers quickly. Thus the objective of this caselet is to draw a general understanding of past initial performances of ex-new consultants, such that Joey's team may have a better gauge of where they stand.

B. METHODOLOGY

We obtained historical data from an ex-agency which was previously successful in recruiting many new agents. We sieved thru the data to present only the ex-new agents' first three months sales performance.

Table 1: Monthly Commissions Based on an Ex-Agency's Performance of New Agents (Altered Data)

Monthly Commissions ($, Based On First 3 Months)	Class Frequencies
6,600 up to 6,800	3
6,800 up to 7,000	7
7,000 up to 7,200	11
7,200 up to 7,400	22
7,400 up to 7,600	45
7,600 up to 7,800	24
7,800 up to 8,000	9
8,000 up to 8,200	4
Total	125

C. HISTOGRAM

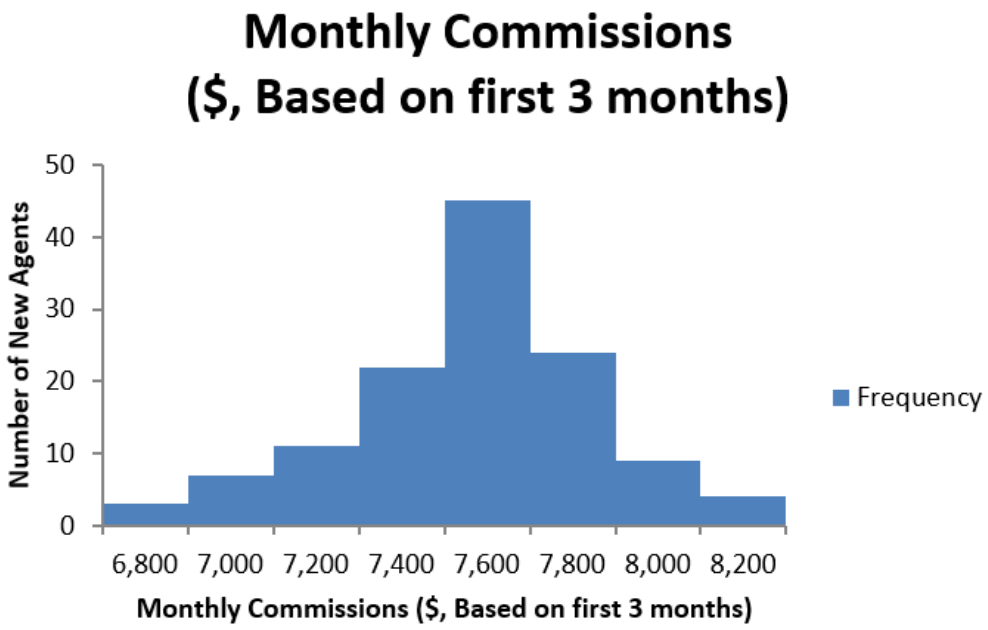


Figure 1: Monthly Commissions of New Agents from Ex-Agency (First 3 Months)

Table 1 shows the data collected for the monthly commissions of new agents from an ex-agency. The data taken was only for the first 3 months of their performance, that is, since day one they stepped into the Financial Advisory (FA) industry. There were a total of 125 new agents being sampled.

Subsequently, we used it to draw a histogram as shown in Figure 1. The steps to draw this histogram can be found in Ang (2020). Its visually evident that the mean (or average) monthly commission is around $7,200 to $7,600. There were 45 new agents who achieved this range of commissions. We also observed that three agents hit a low of between $6,600 to $6,800; while four agents hit a high of $8,000 to $8,200.

We note the open and close brackets [and). For example, [6600, 6800) means inclusive of 6600 but excluding 6800.

D. PIE CHART

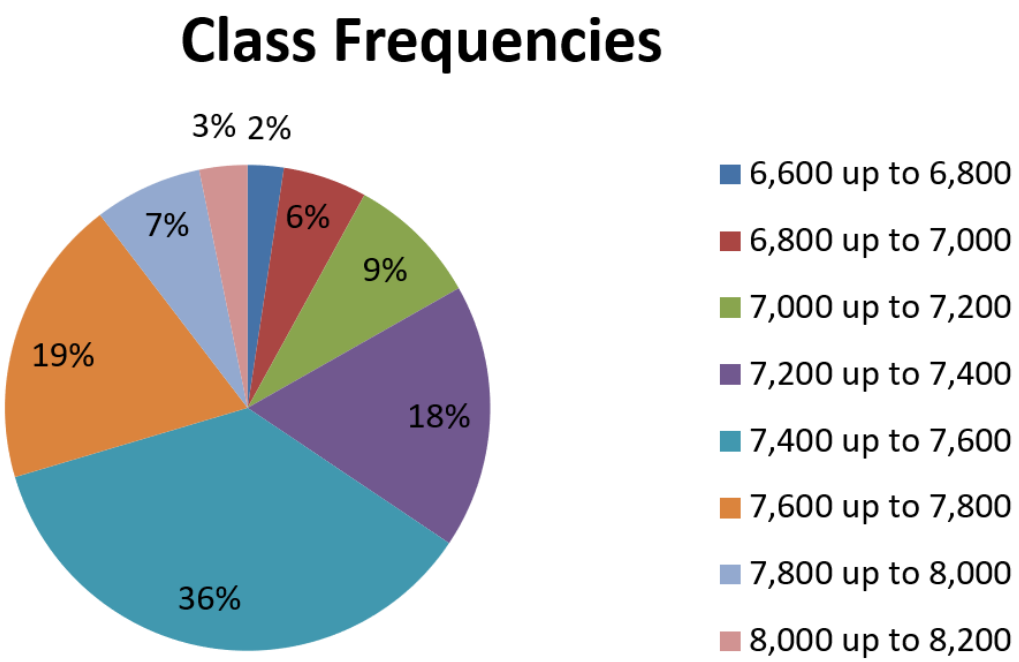


Figure 2: Class Frequencies of Monthly Commissions of New Agents

Figure 2 shows the class frequencies of monthly commissions of the new agents sampled. Likewise, we noticed that the highest percentage presented was in the range of $7,400 to $7,600. The steps to draw this pie chart can be found in Ang (2020).

E. DISCUSSION & INSIGHTS

1. WHAT WERE THE SUCCESS FACTORS THAT LED THE FOUR AGENTS TO QUALIFY FOR THE TOP CATEGORY?

- Naturally, after Figure 1 and Figure 2 were presented, questions arose –
    - What factors influenced the top 4 performing agents?
    - Was it their attitude? Was it their network?
    - And likewise, what led the three worse performing agents to hit the lowest category?
- These questions led to whole new series of possible case studies and investigations, but which we will not dwell into at this preliminary stage.

2. HOW MANY NEW AGENTS SUSTAINED AFTER THE FIRST THREE MONTHS?

- We did not manage to capture data of persistency of these new agents. In other words, we did not know if these agents continued in their career after the first three months (i.e. did they quit).
- Some questions that stemmed here forth were:
    - Does it mean that if a new agent obtained an average commission of above $3,200 per month in his/her first three months, this would infer that he/she you will last (or persevere on) in the industry?
    - Is a period of 3 months too short? Should we collect a minimum of 6 months data before inferring a new agent's persistency?
    - Does a new agent's initial performance solely determine his/her persistency?
- We noted that these questions were rhetorical and non-exhaustive, in which is difficult for us to draw any significant / substantial conclusion at this stage.

- Despite the skimpy results, we still felt a need to applaud ourselves for the effort and spirit put in to come to this level of analytical study – one that is often ignored and lowly placed within agencies.

3. ARE WE ABLE TO RE-LABEL THE PIE CHART FOR EASIER UNDERSTANDING?

- We felt that Figure 2's pie chart could be improved. Thus, we labeled the bins into 3 categories (poor, average, good) for easier reference.

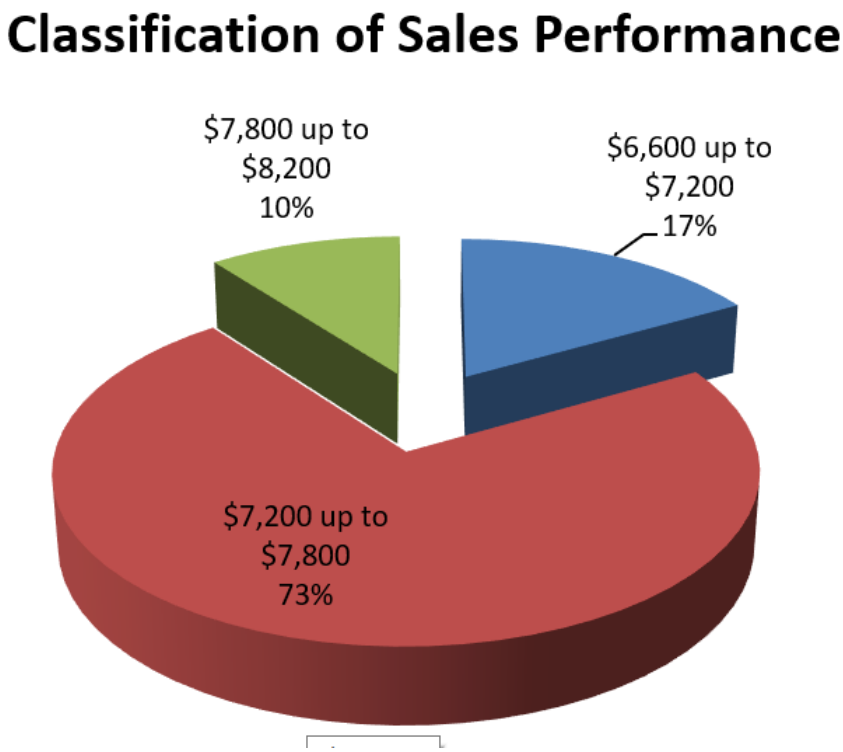


Figure 3: Classification of Sales Performance according to Poor / Average / Good

- Joey's team could now have a bird's eye view of knowing where they stand.
    - Poor Performance:
        - From $6,600 to $7,200 per month.

        - We should expect to see only 17% of new agents in this category.

- Average Performance:

    - From $7,200 to $7,800 per month.

    - We should expect to see 73% of new agents in this category.

- Good Performance:

    - From $7,800 to $8,200 per month.

    - We should expect to see only 10% of new agents qualifying this category.

V. ANALYZING THE PROBABILITY OF ABSENTEESIM OF NEW CONSULTANTS

A. SYNOPSIS

In the initial months of a new consultant in the Financial Advisory (FA) industry, he/she cannot afford to be absent for work. This is because the first few months are crucial to his/her subsequent performance. The first few months will include activities such as training and canvassing, all of which cannot be missed. The objective of this caselet is to gain some gut feel of past absenteeism rates of new consultants from a previous agency.

B. METHODOLOGY

Table 2: Probability of Absenteeism

Number of Days Absent	Probability
0	0.60
1	0.20
2	0.12
3	0.04
4	0.04
5	0.00

Table 2 shows a probability distribution of the likelihood that new consultants (from a previous agency) were absent for 0, 1, 2, 3, 4, or 5 days average in a month. This data was collected from 125 new consultants. We did not question nor doubt the probability obtained / affixed to the number of days absent, rather, we assumed they were just given. We also assumed absenteeism to be "missing in action" or rather, missing from work with a valid reason, for example, having a medical certificate, or having to attend to urgent family / personal matters.

C. DISCUSSION

1. IDENTIFYING THE AVERAGE NUMBER OF DAYS AN EMPLOYEE WAS ABSENT IN A MONTH.

$$E\ x\ = \sum x.P\ x$$

Where:

- E(x): Expected value of x. That is, the mean or average number of days absent in a month, based on the probability table given in Table 2.
- x: Number of days absent in a month of new consultants.
- P(x): Probability of x, number of days absent

Table 3: Calculation of E(x)

Number of Days Absent (x)	Probability (P(x))	x.P(x)
0	0.6	0
1	0.2	0.2
2	0.12	0.24
3	0.04	0.12
4	0.04	0.16
5	0	0
	E(x) =	0.72

Table 3 above shows how E(x) is calculated (using the formula shown above). We identified the average number of days a new consultant was absent in his first 3 months as E(x) = 0.72 days. This meant that on average, a new consultant would be "missing in action" for around ¾ of a day. This was still considered admissible or ok, especially for this previous agency who was successful in grooming new agents (as per their sales results reflected in Table 1).

2. CALCULATING THE STANDARD DEVIATION (SD) OF THE NUMBER OF DAYS ABSENT IN A MONTH.

$$\sigma_x = \sqrt{\sum \left[x - E(x) \right]^2 . P(x)}$$

Where:

- σ_x: Standard Deviation (SD) of the number of days absent in a month.
- E(x): Expected value of x. That is, the mean or average number of days absent in a month, based on the probability table given in Table 2.
- x: Number of days absent in a month of new consultants.
- P(x): Probability of x, number of days absent

Table 4: Calculation of SD(x)

Number of Days Absent (x)	Probability (P(x))	x.P(x)	x-xbar	(x-xbar)^2	(x-xbar)^2.P(x)
0	0.6	0	-0.72	0.5184	0.31104
1	0.2	0.2	0.28	0.0784	0.01568
2	0.12	0.24	1.28	1.6384	0.196608
3	0.04	0.12	2.28	5.1984	0.207936
4	0.04	0.16	3.28	10.7584	0.430336
5	0	0	4.28	18.3184	0
	E(x) = xbar =	0.72		Var(x) =	1.1616
				SD(x) =	1.077775487

Table 4 above shows how SD(x) is calculated (using the formula shown above). We have already identified the average number of days a new consultant was absent in his first 3 months as E(x) = 0.72 days in Table 3. Thereafter, we used E(x) to obtain SD(x), which is approximately around 1.07 days. This meant that on average, a new consultant would be "missing in action" for around ¾ ± 1 of days.

In other words, we may assume a new consultant to be absent (average in a month) for

- 0 days in a good month
- 1 day in a typical month
- 2 days for a bad month.

3. IDENTIFYING THE PROBABILITY THAT A NEW AGENT HAD MORE THAN ONE DAY ABSENT IN A MONTH.

Since we have already identified that a bad month will consist of 2 absent days, and 1 absent day as a typical rate of absenteeism for new consultants, naturally, we would like to discover the probability that a new agent had more than one absent day in a month

$P(x > 1 \text{ day})$

$= P(2) + P(3) + P(4) + P(5)$

$= 0.12 + 0.04 + 0.04 + 0$

$= 0.2$

The calculation above showed us that the probability of a new agent having more than one absent day in a month is 20%. This is pretty low; which we believe is a key reason for the great success of the ex-agency (which we extracted the data from). This also means that if Joey's team is gearing up for success, they would need to monitor their new agents' absenteeism rate – that it should not go beyond a probabilistic 20% in any month.

VI. WILL TRAINING OF NEW CONSULTANTS HELP THEM IMPROVE THEIR RELATIONSHIPS WITH CUSTOMERS?

A. SYNOPSIS & METHODOLOGY

Joey thinks that her team of new consultants will improve their sales figures if they undergo training; specifically, a training program to improve customer relationships. After the training program, we investigated the effectiveness of that training program; whether or not it really improved customer relationships. We did a pre-training and post-training customer survey. A high survey score indicated a good customer relationship. Seven customers were randomly selected and completed both surveys. The survey scores are listed below in Table 5.

Table 5: Pre and Post Training Survey Scores

Customer	Pre-training Survey	Post-training Survey
A	6	8
B	5	5
C	10	10
D	7	10
E	6	8
F	5	6
G	2	8

Table 6: Survey Score Difference of Post - Pre Training

Survey Score Difference of Post - Pre Training	
Mean	**2.00**
Standard Error	0.787
Median	2.000
Mode	2.000
Standard Deviation	2.082
Sample Variance	4.333
Kurtosis	1.807
Skewness	1.242
Range	6.000
Minimum	0.000
Maximum	6.000
Sum	14.000
Count	7.000
Confidence Level (95.0%)	1.925

We apply Descriptive Statistical methods as described in Ang (Descriptive Statistical Measures, 2020) to obtain Table 6. We observed the difference between the pre and post training to be 2.00 points. In other words, the improvement in customer relationship score upgraded only by 2 points (after the new consultants undergone training). This begets the next question: does this mean that the training is effective or otherwise? Is there a significant impact on the customer relationships after they have undergone training? In order to answer this question, we conduct Hypothesis Testing as shown in Ang (Hypothesis Testing, 2020).

Using Ang's (Hypothesis Testing, 2020) Hypothesis Testing, we generate Table 7, a Paired t-Test for two sample means. Purpose here is to determine whether or not the training did significantly impact the consultants.

Table 7: t-Test - Paired Two Sample for Means

t-Test: Paired Two Sample for Means		
	pre-training score	*post-training score*
Mean	5.857	7.857
Variance	5.810	3.476
Observations	7.000	7.000
Pearson Correlation	0.551	
Hypothesized Mean Difference	0.000	
df	6.000	
t Stat	-2.542	
P(T<=t) one-tail	0.022	
t Critical one-tail	1.943	
P(T<=t) two-tail	0.044	
t Critical two-tail	2.447	

First, we will conduct a 2 tailed t test to check if there's a ***significant difference*** between Pre and Post training.

Should there be NO significant difference, we can stop the test (and confirm that the 2 point difference is insignificant).

However, if we found out that there IS a significant difference, naturally, since the Post training score is higher than Pre training, we can confirm that training is effective.

But to double confirm, we will continue with the 1 tailed t test.

1. 2 TAILED TEST

a) Step 1a: Stating the Null and Alternate Hypothesis

$$H_0 : \mu_{Pre} = \mu_{Post}$$
$$H_1 : \mu_{Pre} \neq \mu_{Post}$$

- H0: The Null Hypothesis states that there is NO significant difference between Pre vs Post training score. This also means that the training was effective.
- H1: The Alternate Hypothesis that claims that there IS a significant difference. This means that the training was ineffective.
- μPre : The mean of the Pre training score.
- μPost : The mean of the Post training score.

b) Step 1b: Stating the Level of Significance, a

Normally, a level of significance,

- $\alpha = 10\% \rightarrow$ would mean that we will be 90% confident of our results
- $\alpha = 5\% \rightarrow$ would mean that we will be 95% confident of our results
- $\alpha = 1\% \rightarrow$ would mean that we will be 99% confident of our results

According to Ang (Hypothesis Testing, 2020), the most popular level of significance, $\alpha = 5\%$. Thus, we will stick to this.

c) Step 1c: Choosing the Test Statistic

According to Ang (Hypothesis Testing, 2020), there are three test statistics to choose from:

1. Z test
2. Student's t test
3. F test

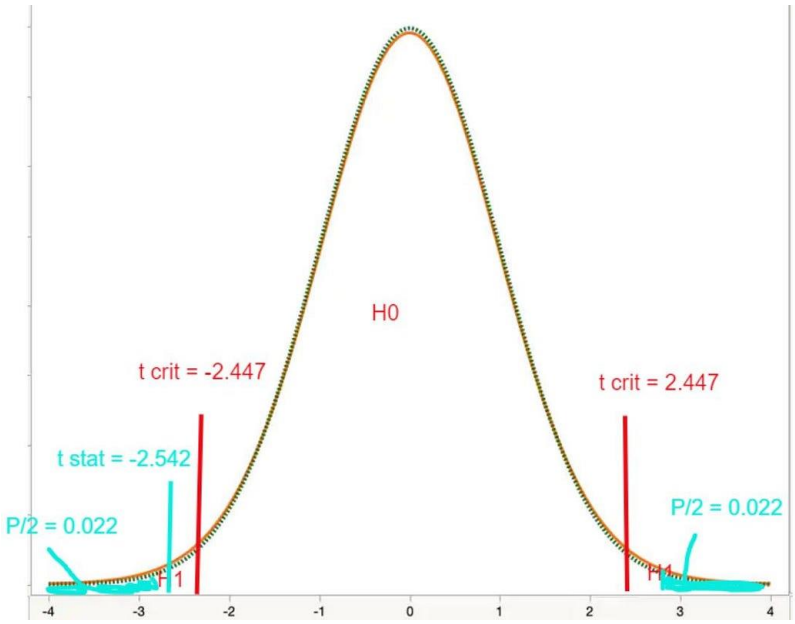
We shall choose the Student's Paired t test because, according to Ang (Hypothesis Testing, 2020),

We do not have the population standard deviation σ (so we can't use the Z test).

We only have two samples — pre vs. post training scores (so we can't use the F test).

The samples are paired, meaning, the seven trainees' scores were taken before (pre) and subsequently, after (post). In other words, they were the same students that were sampled on — thus is a Paired test.

d) Step 1d: Formulating the Decision Rule



In the figure above, we see that t stat falls outside of t critical. At the same time, the P-value (2 tail) = 0.044 < α (0.05) → which means we Accept H1.

e) Step 1e: Making the Decision

Since H1 is accepted, this means that there IS a signifcant difference between Pre vs Post training → Training is effective!

2. 1 TAILED TEST

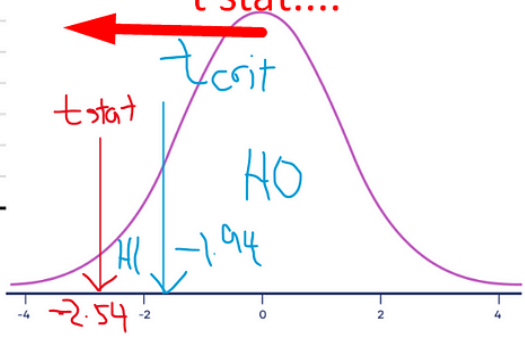
To double confirm our results (even though we can already see that Post training scores are 'significantly' higher than Pre training), we conduct the 1 tailed test.

a) Step 2a: Stating the Null and Alternate Hypothesis

t-Test: Paired Two Sample for Means

	pre-training score	*post-training score*
Mean	5.857	7.857
Variance	5.810	3.476
Observations	7.000	7.000
Pearson Correlation	0.551	
Hypothesized Mean Difference	0.000	
df	6.000	
t Stat	-2.542	
P(T<=t) one-tail	0.022	
t Critical one-tail	1.943	
P(T<=t) two-tail	0.044	
t Critical two-tail	2.447	

here's how we know its LEFT TAIL test... follow the t stat....



$$H_0 : \mu_{Pre} \geq \mu_{Post}$$

$$H_1 : \mu_{Pre} < \mu_{Post}$$

- H0: The Null Hypothesis states that the POST training score is lower than the PRE training score. This also means that the training was ineffective.
- H1: The Alternate Hypothesis states that the POST training score is significantly higher than the PRE training mean score. This also means that the training was effective.

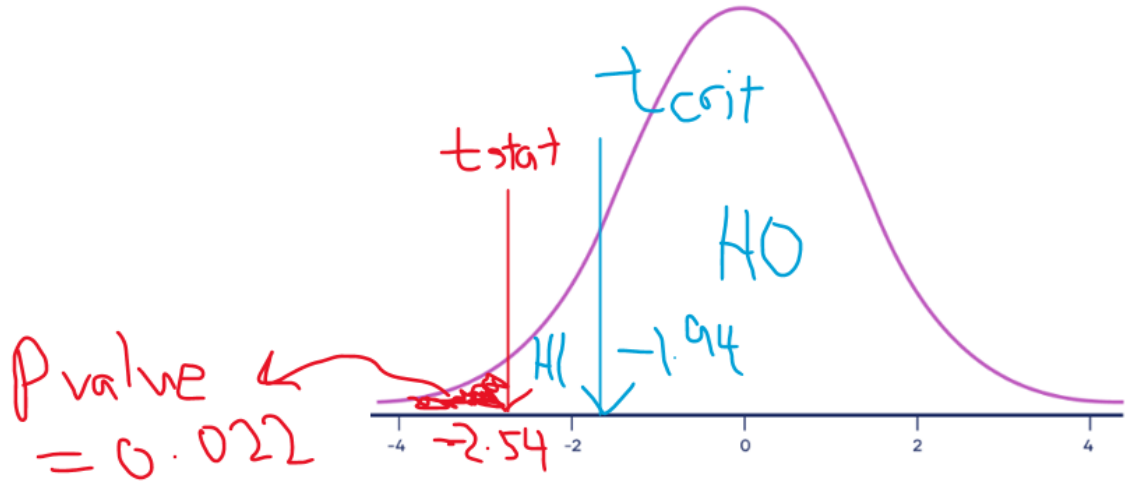
b) *Step 2b: Stating the Level of Significance, a*

According to Ang (Hypothesis Testing, 2020), the most popular level of significance, α = 5%. Thus, we will stick to this.

c) *Step 2c: Choosing the Test Statistic*

We shall choose the Student's Paired t test because these seven trainees' were the same students that were sampled on — thus is a Paired test.

d) *Step 2d: Formulating the Decision Rule*



- The P value (1 tail) = 0.022 < α (0.05) → Accept H1
- Furthermore, the t stat lies in the H1 region.

e) Step 2e: Making the Decision

- Since we Accept H1, meaning μPre < μPost, we have shown once again that Post training scores were higher.

- This concludes that training was effectively significant.

B. DISCUSSION

The belief that training will improve new agents' performance is tested on seven customers. The results showed that training did improve their customer relationships. Of course, one might argue that such a small sample size of seven customers is too small — but that is the best we can do at the moment, due to limited time and budget.

This study helped us gain insights into Joey's teams' training performance. Moving forward, she would most probably be focusing on more customer relationship training.

VII. STUDYING THE FACTORS THAT IMPACT THE SALARIES OF NEW CONSULTANTS

A. SYNOPSIS

We are trying to find out whether or not new agents' previous job experience / gender / and length of service have anything to do with their monthly commissions, or personal sales figures. We managed to obtain past data of 125 new agents from a successful ex-agency. The data consisted of their monthly commissions related to three independent variables:

- Service:
    - length of service, the number of months new agents' have been working in the agency
- Gender:
    - 0 = male, 1 = female
- Previous Job Type:
    - Job type, 0 = white collar worker, 1 = blue collar worker.

Table 8: Some Data for ANOVA Analysis

New Agent	Monthly Commission ($Y)	Service Length (Months)	Gender	Previous Job Type
1	$800.00	2	Male	1
2	$1,256.00	24	Male	1
3	$3,521.00	36	Female	0
4	$299.00	1	Male	0
5	$400.00	8	Female	1

Table 8 shows the data we collated (but only for the first 5 rows). It will then be used for Analysis of Variance (ANOVA) as shown in Table 9.

B. METHODOLOGY

We follow Ang (Analysis Of Variance (ANOVA), 2020) to produce an Analysis Of Variance (ANOVA) table as shown in Table 9.

Table 9: ANOVA Table

ANOVA				
Source of Variation	df	Sum of Squares	Mean Square	F
Regression	3	1004346.771	334782.257	5.96
Residual	26	1461134.596	56197.48445	
Total	29	2465481.367		
	Coefficients	Standard Error	t Stat	p-value
Intercept	784.92	322.25	2.44	0.02
Service	9.19	3.20	2.87	0.01
Gender	222.78	89.00	2.50	0.02
Job	-28.21	89.61	-0.31	0.76

1. RELATIONSHIP BETWEEN MONTHLY COMMISSION, SERVICE LENGTH, GENDER AND PREVIOUS JOB TYPE

$$Y = \$784.92 + \$9.19\ X1 + \$222.78\ X2 - \$28.21\ X3$$

Where:

- Y: New Agent's Monthly Commission ($)
- X1: Length of Service (Months)
- X2: Gender (0 = male, 1 = female)
- X3: Previous Job Type (0 = white collar worker, 1 = blue collar worker)

The above equation shows us the relationship between Y and the independent variables. An example could be:

- For a start, if X1, X2, X3 = 0
    - meaning the agent has just started with zero work experience (X1 = 0)
    - is male (X2 = 0)
    - was previously a white collared worker (X3 = 0)
- Then the initial commission is $784.92 per month. This is roughly around $785. But one will question: how is it possible for an entirely new agent to achieve $785 if she hasn't sold any policies yet? This is because we have reason to believe that the ex-agency had sponsorship. Meaning, they gave some allowance for entirely new agents, as a means to encourage them during the initial tough phase.
- Another example could be if X1, X2, X3 = 1
    - meaning the agent has just started with 1 month work experience (X1 = 1)
    - is female (X2 = 1)
    - was previously a blue collared worker (X3 = 1)
    - Y = $784.92 + $9.19 + $222.78 - $28.21 = $988.68 per month
    - We expect this new female agent to be making around $988 per month.
- Thus,
    - Every increase in X1 (length of service in months) by 1 month → Y will increase by $9.19
    - If X2 (0 = male, 1 = female) is a female → Y will increase by $222.78
    - If X3 (0 = white collar worker, 1 = blue collar worker) is a blue collared worker → Y will decrease by $28.21

a) How fitting is the Multiple Regression line?

Though we now know the Multiple Regression equation representing the relationship between the variables (Y = $784.92 + $9.19 X1 + $222.78 X2 - $28.21 X3), we would like to find out how fitting are the variables. Meaning, how close are they to a straight line regressed to Y (the monthly commissions).

The indicator used here is called the Coefficient of Multiple Determination, R^2 (Ang, Multiple Regression, 2020).

Coefficient of Multiple Determination $R^2 = \frac{SSR}{SS\ total}$

We see from Table 9 that the Regression Sum of Squares (SSR) = 1004346; while the Total Sum of Squares (SS total) = 2465481. Thus, R^2 can be calculated as (1004346.771/2465481.367) = 0.407 or 40.7 percent.

R^2 must always be between 0 and 1, inclusive. That is, $0 \leqslant R^2 \leqslant 1$. • The closer R^2 is to 1.0, the stronger the association between Y and the set of independent variables, X1, X2, X3. For example, if R^2 = 0.407 for the Y hat equation given above, that means that X1, X2 and X3 account for 40.7 percent of the variation of Y (Ang, Multiple Regression, 2020).

40.7% means that the regression line is very weak. A strong regression line would be around 90%. In other words, we might not be sure if the variables X1 (length of service), X2 (male / female) and X3 (blue or white collared job) are really the cause of Y (monthly commissions), but this is the best data we are presented with at the moment.

In order to try and resolve this, we carry out the next step: checking out which variables are significant / insignificant. This is so that we can eliminate the unimportant ones.

2. WHICH ARE THE SIGNIFICANT / IMPORTANT VARIABLES?

a) Global F test

We execute a Global F hypothesis test to determine whether all regression coefficients are zero at the 5% significance level. In other words, we are trying to find out whether X1, X2 and X3 are even important at all; do they affect Y, or if not, the entire model can be scrapped.

(1) Step 1: Stating the Null and Alternate Hypothesis

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0$$
$$H_1: \text{Not all the } \beta s = 0$$

Interpretation: IF H0 is accepted, it implies that ALL of the independent variables, X1, X2, etc.., is INSIGNIFICANT in causing changes in Y hat.
IF H1 is accepted, this means that ONE OR MORE of the independent variables, X1, X2, etc.. is SIGNIFICANT in causing changes in Y hat. Note that it does not suggest how many, or which of the independent variables are significant.

Where:

- H_0: The Null Hypothesis, which states that all variables (X1, X2, X3) are unimportant and do not affect Y at all. If this occurs, it means that the entire Multiple Regression model can be scrapped – because it is not futile to derive any insights from this.

- H_1: The Alternate Hypothesis, which states that at least one of the variable (X1, X2, X3) is important and does affect Y. If this occurs, we will proceed to the next phase: Individual F test to check which are the important variables.

(2) Step 2: Stating the Level of Significance, α

We shall choose an α of 0.05 because it is the most common as stated by Ang (Hypothesis Testing, 2020).

(3) Step 3: Choosing the Test Statistic

Since we are doing an Analysis of Variance (or ANOVA test), this is the Global F test → the F statistic is chosen.

(4) Step 4: Forming the Decision Criteria

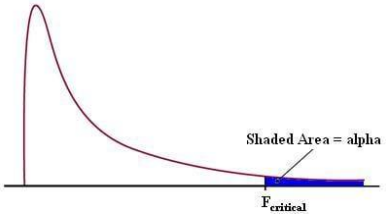
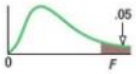


Figure 4: F Distribution Decision Criteria (Ang, Multiple Regression, 2020)



Degrees of Freedom for the Denominator	Degrees of Freedom for the Numerator															
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39

Figure 5: F Distribution: Locating F Critical

Figure 4 shows the F Distribution Decision Criteria. The x axis represents the values of F. Anything before F Critical, we accept H0 (and reject H1). Anything after F Critical, we accept H1 (and reject H0).

Figure 5 shows how we locate the F Critical through the F distribution table (for $\alpha = 5\%$). We see from Table 9 that the F Degree of Freedom (df) of the Regression is 3 (numerator) while that of the Residual (denominator) is 26. We look at Figure 5 and the closest value that it gives is 2.99.

(5) Step 5: Making the Decision

That is, F Critical = 2.99 while F Statistic (from Table 9) = 5.96. Since F Critical > F Statistic, it means that the decision has fallen into the H1 region. So we accept H1 and reject H0.

(6) Step 6: Conclusion

From this Global F test, we see that H1 is accepted, that is to say that not all the variables are redundant. In other words, this Multiple Regression (MR) model still holds value (and should not be scrapped, as we thought of earlier in the previous section "How fitting is the Multiple Regression line?"). Thus, even though the MR model doesn't fit well (since the R^2 is very low), the model still holds.

So we know currently that a combination of X1 (length of service in months), X2 (male or femal) and X3 (white or blue collared previous job) does affect Y (the monthly commission). But which of them are the more important variables? We find out in the next section when we perform Individual F test.

b) Individual F test

We execute a Individual F test to determine which regression coefficients are zero at the 5% significance level. In other words, we are trying to find out which variables: X1, X2 or X3 are important; and which ones can be dropped.

(1) Step 1: Stating the Null and Alternate Hypothesis

X1: Length of Service (months)	X2: Gender (male / female)	X3: Previous Job Type (blue / white collared job)
$H_0 : \beta_1 = 0$ $H_1 : \beta_1 \neq 0$	$H_0 : \beta_2 = 0$ $H_1 : \beta_2 \neq 0$	$H_0 : \beta_3 = 0$ $H_1 : \beta_3 \neq 0$

Where:

- β1 is the residual for X1 (and so forth)
- H_0: The Null Hypothesis, which states that the residual (β1 for X1, and so forth) is unimportant and does not affect Y at all. If this occurs, it means that the particular X can be dropped off the Multiple Regression model – because it is not important and does not affect Y much.
- H_1: The Alternate Hypothesis, which states that the residual (β1 for X1, and so forth) is important and does affect Y. If this occurs, we will retain the variable because it significantly impacts Y.

(2) Step 2: Stating the Level of Significance, α

We shall choose an α of 0.05 because it is the most common as stated by Ang (Hypothesis Testing, 2020).

(3) Step 3: Choosing the Test Statistic

Since we are doing an Analysis of Variance (or ANOVA test), this is the Individual F test → the F statistic is chosen.

(4) Step 4: Forming the Decision Criteria

Figure 4 shows the F Distribution Decision Criteria. The x axis represents the values of F. Anything before F Critical, we accept H0 (and reject H1). Anything after F Critical, we accept H1 (and reject H0).

The shaded region as represented in Figure 4 is the Level of Significance, $\alpha = 5\%$). Should the p-value exceed 0.05, it would have spread across the region, consuming into H0; and vice versa.

(5) Step 5: Making the Decision

We see from Table 9 that the P-values of:

- $\beta 1$: p-value = 0.01 < α = 0.05 → Accept H1
- $\beta 2$: p-value = 0.02 < α = 0.05 → Accept H1
- $\beta 3$: p-value = 0.76 > α = 0.05 → Accept H0

This means that $\beta 3$ (or rather, X3) is unimportant; and should be dropped off. Whereas $\beta 1$ and $\beta 2$ (or rather, X1 and X2) are important and should be kept.

(6) Step 6: Conclusion

From this Individual F test, we see that H1 is accepted for both X1 and X2, that is to say that the length of service and the gender does significantly impact the monthly commissions. However, H0 is accepted for X3; which means that the previous job type (blue or white collared worker) does not impact the monthly commissions.

C. DISCUSSION

From this study, Ms. Joey is now better able to understand the factors that impact a new agent's monthly performance. It isn't about his/her past job type that is important; rather, the focus should be on the length of service and gender type. This will help Joey in future recruitment planning. For example, when recruiting new agents, she might be looking out for important factors such as:

- Is the new applicant experienced in the financial consultancy industry? I.e. has he/she been an agent before? If so, for how long?

- This is because the more experienced an agent, Joey will predict that he/she is likely to achieve higher commissions.

- Is the agent male or female? If she's female, the likelihood of her performing better than her male counterparts is higher.

VIII. AMOUNT OF INSURANCE PREMIUM TO CHARGE

A. SYNOPSIS

Insurance underwriting relies heavily on statistics to determine the amount of insurance premium to charge. However, underwriting is the job scope of the underwriter – not a financial consultant. Most of the time, financial planners / advisors are issued with preset documents to present to their clients. Pricing of products are done backend.

Although it is not the responsibility of the financial advisor to know how products are priced, a little knowledge won't hurt. Besides, they may appear more intelligent if and when their clients ask them about product pricings – the agents may have some basic idea of how they were derived.

B. METHODOLOGY

In this example, we discuss about a term life insurance policy for a 25 year old female. The probability that a 25 year old female will live another year is 0.99786 based on data from an ex - national registry agency. We show here how we calculate the insurance premium an insurance company would charge to break even on a 1 year $0.5 million term life insurance policy.

- Probability that the client will live = 0.99786
- Probability that the client will not live = 1 – 0.99786 = 0.00214

$$E\ x\ = \sum x.P\ x$$

Where:

- E(x): Expected value of x.
- x: 0 = die; 1 = live.
- P(x): Probability of x

- E (X) = ($0)(0.99786) + ($500,000)(0.00214) = $1,070

The insurance premium that the company should charge (per year) based on a 1 year term life policy for the 25 year old females is $1,070.

C. DISCUSSION

In this short example, we have shown how an actuaries / underwriter calculate the yearly premium for a 25 year old female. It does help give some idea to new advisors about the kind of work the backend people do in insurance companies.

IX. COMPARING THE YIELD OF INCOME V.S. GROWTH ORIENTED MUTUAL FUNDS

A. SYNOPSIS

Many a times in real life, clients ask their advisors to do product comparisons. Especially when it comes to investments, they want their advisor to do due diligence in squeezing every penny. In this example, we assume that the client has tasked his financial planner to compare the yield of income vs. growth oriented mutual funds. In other words, he wants to find out: is there a difference in the mean yield of these two funds?

The client presented a fictitious amount of hundred thousand dollars to invest – fifty thousand in each of a sample of 35 income oriented and 40 growth oriented funds. For a period of two years,

- The mean increase (yield) for the Income Fund (IF) was $1,100.
    - The Standard Deviation of the IF is σ_I = $45
- The mean increase (yield) for the Growth Fund (GF) was $1,090.
    - The Standard Deviation of the GF is σ_F = $55

B. METHODOLOGY

1. STEP 1: STATING THE NULL AND ALTERNATE HYPOTHESIS

$$H_0 : \mu_i = \mu_g$$
$$H_1 : \mu_i \neq \mu_g$$

Where:

- μ_i: Mean yield (increase) of the Income Fund (IF)
- μ_g: Mean yield (increase) of the Growth Fund (GF)
- H_0: The Null Hypothesis, which states that the mean yield (increase) of the Income Fund (IF) is equal to the mean yield (increase) of the Growth Fund (GF). If this occurs, it means that the client can allocate his funds into either fund, yet experience the same yield. Put simply, it does not matter which fund he puts in: IF or GF.
- H_1: The Alternate Hypothesis, which states that the mean yield (increase) of the Income Fund (IF) is not equal to the mean yield (increase) of the Growth Fund (GF). If this occurs, it means that the client needs to probe further – which fund IF or GF has a greater yield? And are there any other aspects that need to be considered before investing? He could possibly ask his advisor to do more homework on this.

2. STEP 2: STATING THE LEVEL OF SIGNIFICANCE, A

We shall choose an α of 0.01, so that we can be 99% confident of our results.

3. STEP 3: CHOOSING THE TEST STATISTIC

Test Statistic for No Difference Between Two Sample Means

$$z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Where: $\overline{X}_1$ and $\overline{X}_2$ refer to the two sample means.
σ_1^2 and σ_2^2 refer to the two sample variances.
n_1 and n_2, refer to the two sample sizes.

Figure 6: Test Statistic for No Difference between Two Sample Means (Ang, Hypothesis Testing, 2020)

We are performing a test for No Difference between Two Sample Means (as can be seen in the equation above as extracted from (Ang, Hypothesis Testing, 2020)). We choose the Z test statistic because:

- The two samples are independent (IF is not related to GF in any way)
- Both their Population Standard Deviations are known (σ_I = \$45; σ_F = \$55)

4. STEP 4: FORMING THE DECISION CRITERIA

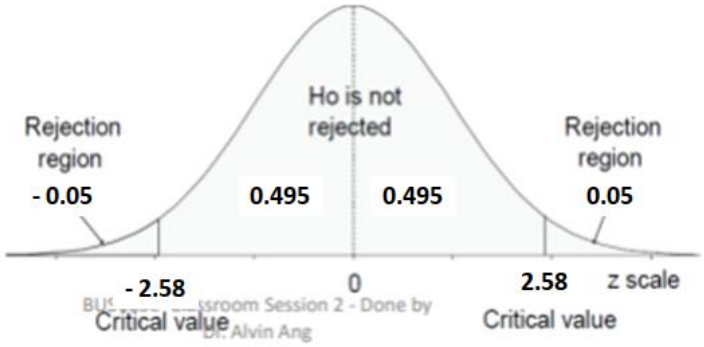


Figure 7: Forming the Decision Criteria using the Z Distribution (Ang, Hypothesis Testing, 2020)

Figure 7 shows the decision criteria. We reject H0 if z < -2.58 or z > 2.58. In other words, the z critical = ± 2.58.

5. STEP 5: MAKING THE DECISION

$$z = \frac{1100 - 1090}{\sqrt{\frac{(45)^2}{35} + \frac{(55)^2}{40}}} = 0.87$$

We apply the equation as shown in Figure 7 and we obtain z statistic = 0.87. In other words, the z statistic falls within the range of the z critical

- $-2.58 < 0.87 < 2.58$

This means that H0 is not rejected.

6. STEP 6: CONCLUSION

From this Z test, we see that there is no difference in the yields of either IF or GF. It means that the client can allocate his funds into either fund, yet experience the same yield. Put simply, it does not matter which fund he puts in: IF or GF.

X. HOW DOES GENDER AND TYPE OF INSURANCE POLICY AFFECT THE NUMBER OF POLICIES SOLD?

A. SYNOPSIS

For Hospitalization Insurance in Singapore, there are basically three types of Integrated Shield Plans (IPs): Plan A, B or C. These are covered by Private Insurers (shown below).

AIA HealthShield Gold Max

HSBC Life Shield

GREAT SupremeHealth

Singlife Shield

IncomeShield

PRUShield

Raffles Shield

To find our more about what IPs are, you may refer here: https://www.moh.gov.sg/healthcare-schemes-subsidies/medishield-life/about-integrated-shield-plans

An example of what a typical IP will cover, you may take a look at AIA's HealthShield Gold, in their brochure here:

https://www.aia.com.sg/content/dam/sg-wise/en/docs/our-products/health/aia-health-shield-gold-max-english-brochure.pdf

Joey has vast experience in selling these plans and the table below shows the number of IPs she sells in a typical month.

Health Shield Plan	A	B	C
Male	8	10	8
	4	8	6
	0	6	4
Female	14	4	15
	10	2	12
	6	0	9

She wants to determine (from her past selling record above):

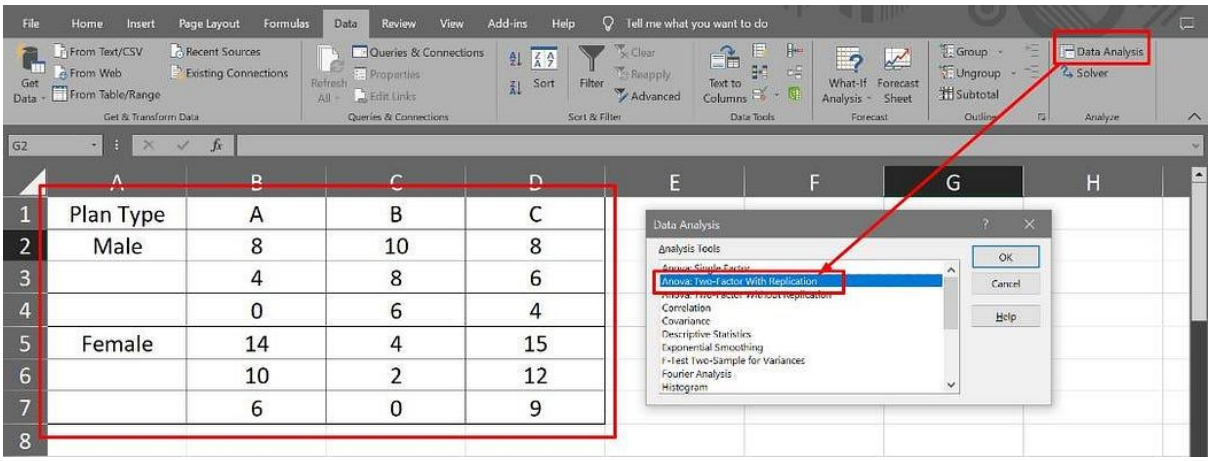
- Does Gender impact buying behaviour?
- Does the Plan Type affect sales?
- Is there an Interaction / Correlation between Gender and Plan Type?

Just by eyeballing the small dataset above, she has already come up with a few preliminary conclusions:

1. It appears that her female customers buy more Shield Plans compared to male ones.
2. It looks as if female customers are more inclined towards buying Plans A and C while males prefer Plan B.
3. For males, they don't seem to bother much about the difference in Plan Types.

Thus we set out to uncover whether her preliminary observations are correct.

B. TWO WAY ANOVA IN EXCEL



	A	B	C	D
1	Plan Type	A	B	C
2	Male	8	10	8
3		4	8	6
4		0	6	4
5	Female	14	4	15
6		10	2	12
7		6	0	9

We shall be using Two Way Anova technique here in Excel. We choose the one with Replication because for each Gender type, we have 3 samples each. That is, for Males, we have 3 rows of data, and likewise for Females.

Anova: Two-Factor With Replication						
SUMMARY	A	B	C	Total		
Male						
Count	3	3	3	9		
Sum	12	24	18	54		
Average	4	8	6	6		
Variance	16	4	4	9		
Female						
Count	3	3	3	9		
Sum	30	6	36	72		
Average	10	2	12	8		
Variance	16	4	9	28.25		
Total						
Count	6	6	6			
Sum	42	30	54			
Average	7	5	9			
Variance	23.6	14	16			
ANOVA						
ource of Variati	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	18	1	18	2.037735849	0.178939877	4.747225347
Columns	48	2	24	2.716981132	0.10634348	3.885293835
Interaction	144	2	72	8.150943396	0.005810254	3.885293835
Within	106	12	8.833333333			
Total	316	17				

The report churned out for our Two Way Anova.

The basis of our Hypothesis Testing will be:

- H0: The means of both Gender groups are equal.
- H1: The means of both Gender groups are different.
- H0: The means of all Plan Types are equal.

- H1: The mean of at least one Plan Type is different.
- H0: There is no interaction between Gender groups and Plan Types
- H1: There is interaction between Gender groups and Plan Types

C. IS THERE A SIGNIFICANT DIFFERENCE BETWEEN SELLING TO A MALE VS FEMALE?

ANOVA						
ource of Variati	SS	df	MS	F	P-value	F crit
Sample	18	1	18	2.037735849	0.178939877	4.747225347
Columns	48	2	24	2.716981132	0.10634348	3.885293835
Interaction	144	2	72	8.150943396	0.005810254	3.885293835
Within	106	12	8.833333333			
Total	316	17				

Sample (or rows) refer to Males vs Females. Its P value is 0.17 (which is > a (0.05)).

Thus, we ACCEPT H0 → There is no difference in sales between Males or Females.

Anova: Two-Factor With Replication				
SUMMARY	A	B	C	Total
Male				
Count	3	3	3	9
Sum	12	24	18	54
Average	4	8	6	6
Variance	16	4	4	9
Female				
Count	3	3	3	9
Sum	30	6	36	72
Average	10	2	12	8
Variance	16	4	9	28.25

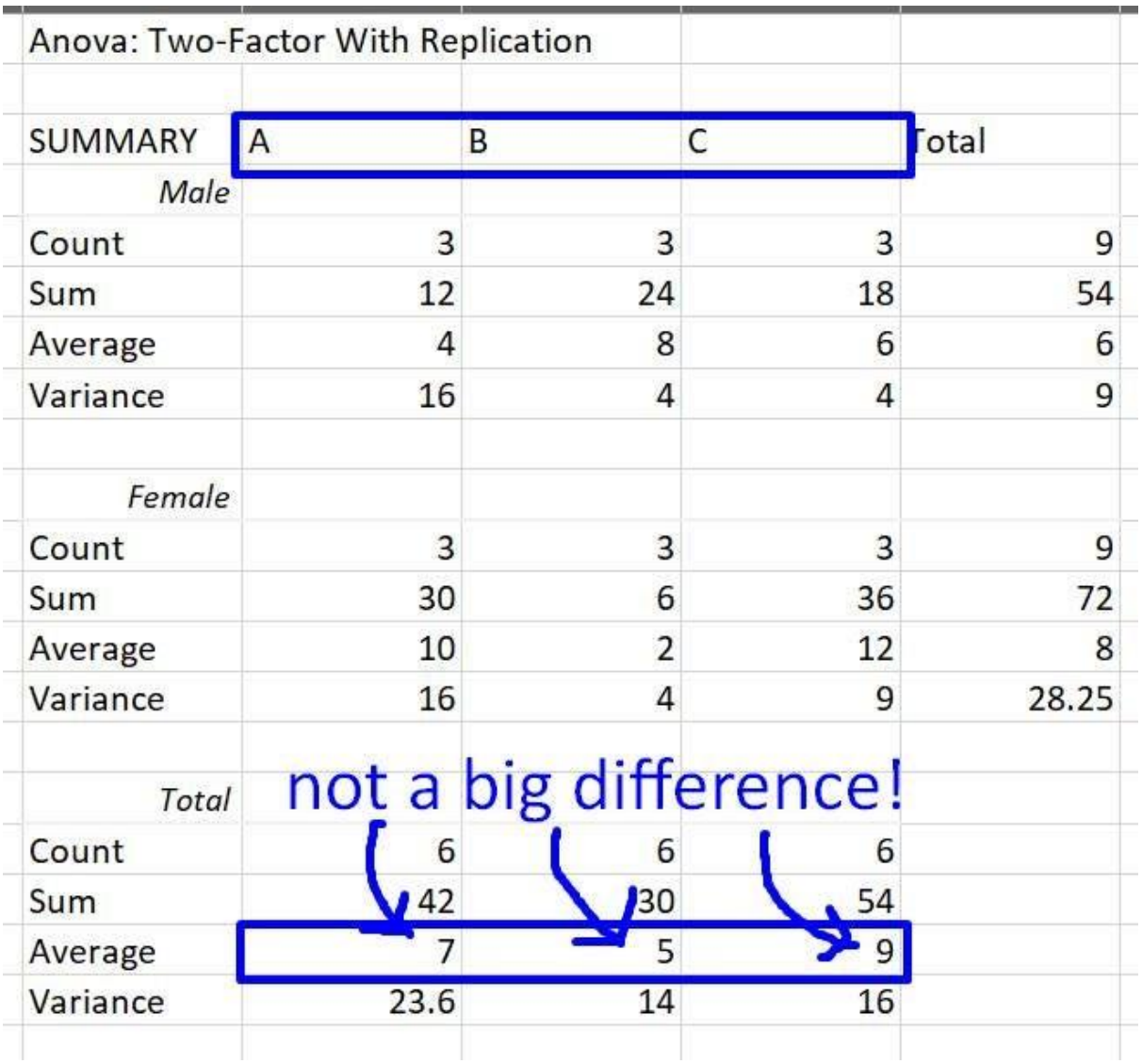
It does not appear that there is a similar amount of sales from Males and Females. Gender doesn't make a difference.

D. IS THERE A SIGNIFICANT DIFFERENCE IN SALES AMOUNT BETWEEN POLICIES TYPE A, B OR C?

ANOVA						
ource of Variati	SS	df	MS	F	P-value	F crit
Sample	18	1	18	2.037735849	0.178939877	4.747225347
Columns	48	2	24	2.716981132	0.10634348	3.885293835
Interaction	144	2	72	8.150943396	0.005810254	3.885293835
Within	106	12	8.833333333			
Total	316	17				

Columns refer to Plan Types. Its P values is 0.106 (which is > a (0.05)).

Thus, we Accept H0 → There is no difference between the sales of A, B or C.



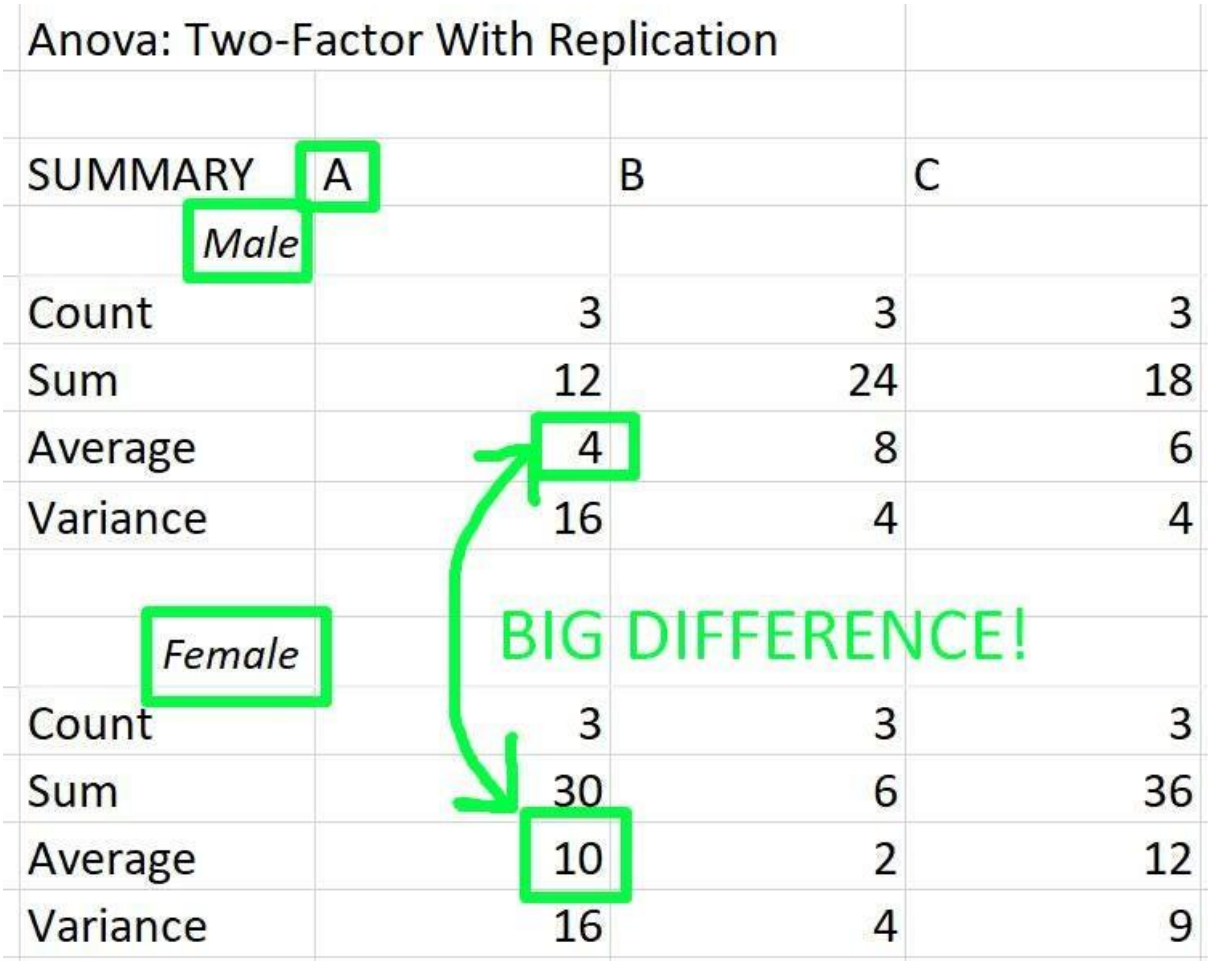
The amount of sales for A, B and C is similar.

E. IS THERE A RELATIONSHIP / INTERACTION BETWEEN GENDER AND PLAN TYPE?

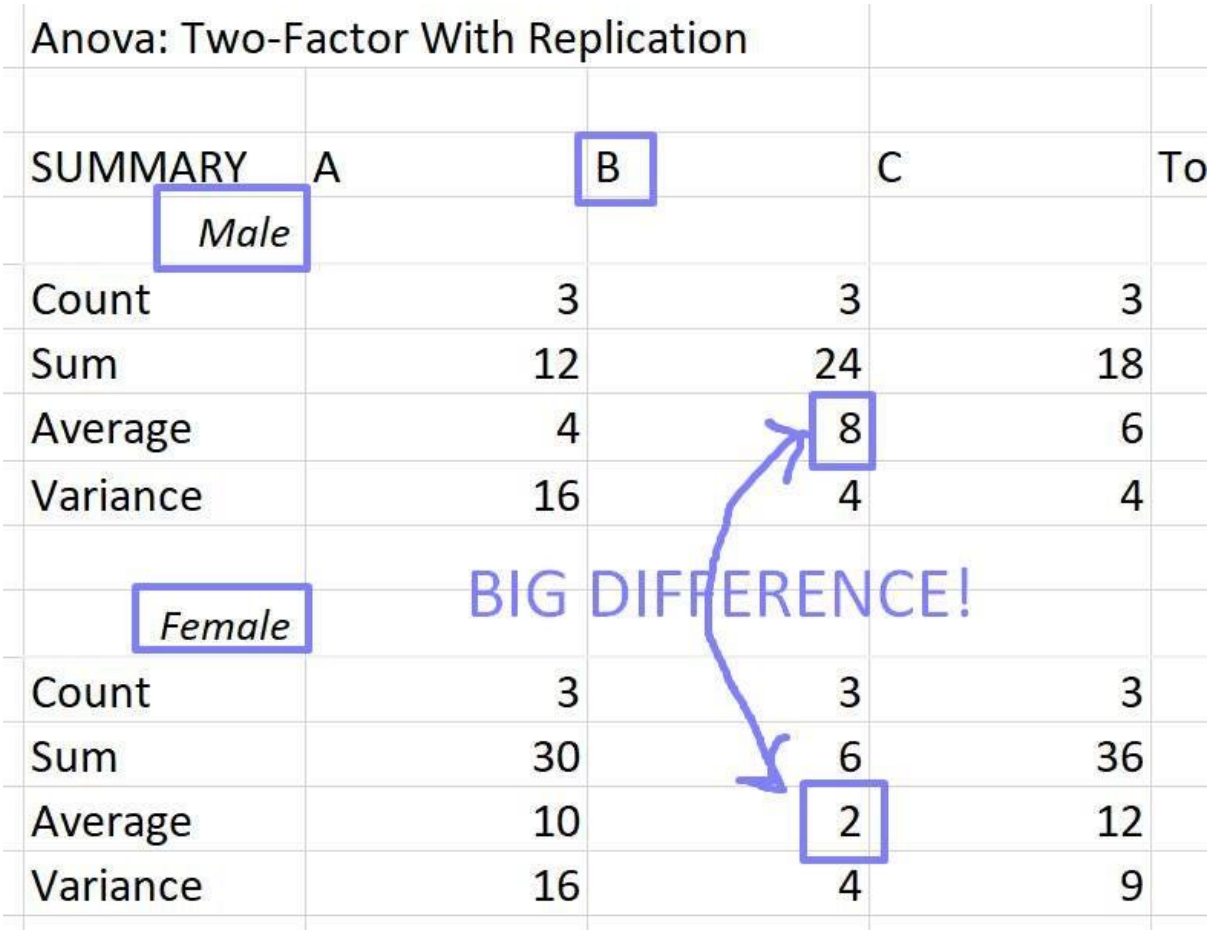
ANOVA						
ource of Variati	SS	df	MS	F	P-value	F crit
Sample	18	1	18	2.037735849	0.178939877	4.747225347
Columns	48	2	24	2.716981132	0.10634348	3.885293835
Interaction	144	2	72	8.150943396	0.005810254	3.885293835
Within	106	12	8.833333333			
Total	316	17				

The interaction effect P-value is .0058 (which is < a (0.05)).

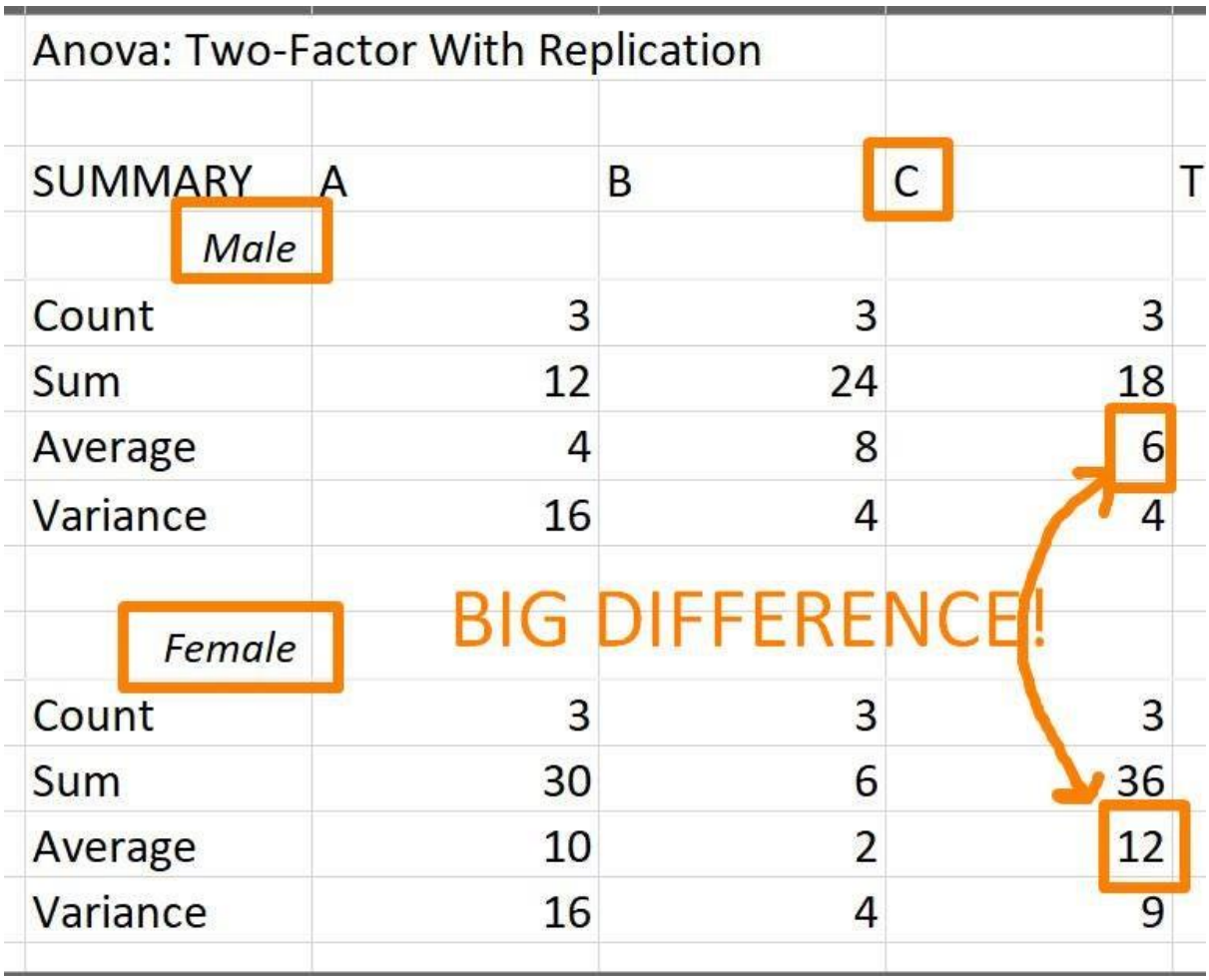
Thus, we Accept H1 → There is an Interaction between Gender and Plan Type.



Seems like Females prefer buying Plan Type A as compared to Males.



Seem like Males prefer buying Plan Type B as compared to Females.



Seems like Females prefer buying Plan Type C over Males.

F. CONCLUSION

1. DOES GENDER IMPACT BUYING BEHAVIOUR?

As whole → No, there's no significant difference in Sales between Males vs Females.

Moving forward, in general, this could potentially mean that Joey can expect her agents to sell roughly the same number of Shield Plans to both Male and Female clients.

2. DOES THE PLAN TYPE AFFECT SALES?

As a whole → No, there's no significant difference in Sales between the different types of plans.

This could mean that in future, her agents can assume that the number of Shield Plan sales they bring in will be roughly equal.

3. IS THERE AN INTERACTION / CORRELATION BETWEEN GENDER AND PLAN TYPE?

Yes. It makes a difference. When Gender and Plan Types are combined, we see the effect on number of Sales.

Her agents need to take note that, while approaching:

Female Prospects, they tend to favour Plan Types A and C.

Male Prospects, they tend to favour Plan Types B.

XI. BIBLIOGRAPHY

Ang, A. (2020). *Analysis Of Variance (ANOVA)*. Retrieved from www.AlvinAng.sg: https://www.alvinang.sg/s/ANOVA-by-Dr-Alvin-Ang-watermarked.pdf

Ang, A. (2020). *Descriptive Statistical Measures*. Retrieved from www.AlvinAng.sg: https://www.alvinang.sg/s/Descriptive-Statistical-Measures-by-Dr-Alvin-Ang-watermarked-prtt.pdf

Ang, A. (2020). *Hypothesis Testing*. Retrieved from www.AlvinAng.sg: https://www.alvinang.sg/s/Hypothesis-Testing-by-Dr-Alvin-Ang-watermarked.pdf

Ang, A. (2020). *Multiple Regression*. Retrieved from www.AlvinAng.sg: https://www.alvinang.sg/s/Multiple-Regression-MR-by-Dr-Alvin-Ang.pdf

XII. ABOUT THE AUTHORS

A. ABOUT JOEY WU

Joey is a Certified Financial Planner (CFP), Institute of Banking and Finance Advanced Specialist (IBFA) and an Associate Estate Planning Practitioner (AEPP). She has vast experience in the Financial Advisory Industry and has also previously qualified for the Million Dollar Round Table (MDRT) and Court of the Table (COT).

More about her at http://joeycfp.weebly.com

B. ABOUT DR. ALVIN ANG

Dr. Alvin Ang earned his Ph.D., Masters and Bachelor degrees from NTU, Singapore. He was a previously a Professor, Scientist and Financial Consultant. Currently, he owns multiple self-started businesses and is a Personal/Business Advisor as well.

More about him at www.AlvinAng.sg