I

Does a raise in yearly increments improve tenures of Google software engineers?

April 15, 2021

MARCUS LOKE, ANNI YI, YUSEN ZHANG, ROMAULI BUTARBUTAR | RESEARCH DESIGN PROJECT PRESENTATION

Today's agenda

Presentation flow

- Introduction and problem statement
- Research questions and hypotheses
- Research plan
- Simulation results

13%

In 2018, the tech sector suffered the highest turnover rates of 13.2% as compared to other industries.

462/500

Google stands at 462nd spot out of Fortune 500 companies for its low employee retention rate.

43%

According to a Deloitte survey, close to half of respondents chose "pay" as the top reason for leaving a company.

71%

According to The Dice salary report, 71% cited "seeking salary compensation" as the top reason for leaving a company.

5

REPERCUSSIONS OF ATTRITION

Slow the business and productivity losses

If a software engineer leaves, it takes 43 days on average to hire a new one (approx. 1.5 months of productivity loss).

Revenue loss

Cost around $33K for each employee who leaves.

Loss of intellectual capital

Create bottlenecks and reduces morale of the team.

2.7% YEARLY INCREMENT

According to the US Bureau of Labor Statistics (2021 report), yearly increments for tech firms was 2.7%, which begs the question...

RESEARCH QUESTION

Can a salary increment of 5% at the end of the first year increase the average tenure for Google software engineers in the United States?

HYPOTHESES

01

NULL HYPOTHESIS

A salary increment of 5% at the end of the first year **does not** increase the average tenure of Google software engineers.

$$H_0 : T_{avg,5\%} - T_{avg,2.7\%} \leqslant 0,$$

02

ALTERNATIVE HYPOTHESIS

A salary increment of 5% at the end of the first year **improves** the average tenure of Google software engineers.

$$H_1 : T_{avg,5\%} - T_{avg,2.7\%} > 0$$

RESEARCH PLAN

1

POPULATION OF INTEREST

- 160 Google software engineers
- 6 months < Tenure < 1 year

2

SAMPLE SELECTION

- Cluster & random sampling across 4 different city offices
- Exclude those with poor evaluations
- Exclude those that Google's not inclined to retain

3

COMPARISON

- Treatment group (5% increment)
- Control group (2.7% increment)

Treatment Groups (5% Increment)	Control Groups (2.7% Increment)
San Francisco, CA, Office (20 people)	San Francisco, CA, Office (20 people)
New York City, NY, Office (20 people)	New York City, NY, Office (20 people)
Sunnyvale, CA, Office (20 people)	Sunnyvale, CA, Office (20 people)
Chicago, IL, Office (20 people)	Chicago, IL, Office (20 people)

"Why a treatment/control group in each city office?"

In order to mitigate the influence of potential confounding factors or multicollinearity due to the different standards of living between cities.

RESEARCH PLAN

4

VARIABLES

- Independent variable: The salary increment of either 5% (treatment) or 2.7% (control).
- Dependent variable: The tenure of employees measured in years

5

STATISTICAL ANALYSIS PLAN

- Two-sample t-test with one-sided alternative ("greater")
- To test the difference in mean tenures between the treatment and control groups

6

DATA COLLECTION

- Work with Google HR to obtain list of employees who satisfy inclusion criteria
- Randomize based on last name
- Need clearance from Google Management

REDUCE RISK OF INTERMINGLING

To reduce risk of intermingling, we will find divisions that are relatively distinct for the control and treatment groups.

E.g., control group from Google Play and treatment group from Google Cloud.

LIMITATIONS & UNCERTAINTIES

1	2
Unmeasured variables may influence the dependent variable (e.g., family income/wealth, employee satisfaction, yearly bonus, staff benefits, etc.)	Intermingling/sharing of salary may still be possible (e.g., subjects of treatment group may be friends with subjects of control group)

SIMULATION RESULTS

14

BACKGROUND

We select four city offices as our experimental subjects. In each city there will be two groups of participants – treatment group and control group; each group has 20 participants. Then we compare the difference in mean tenure.

Sample size

160

Effect size

0.5 years

Confidence interval

95%

Power

90%

STANDARD DEVIATION

Based on initial assumptions, we compute
SD = effect size / d = 1.075955

```
pwr.t.test(n=80, sig.level=0.05, power=0.9, type="two.sample", alternative="greater")
```

```

     Two-sample t test power calculation 

              n = 80
              d = 0.4647034
      sig.level = 0.05
          power = 0.9
    alternative = greater

NOTE: n is number in *each* group
```

METHODOLOGY

EFFECT OF 0.5 YEARS

01

ONE TIME EXPERIMENT

Control: mean = 1.9 years (based on literature review)
Treatment: mean = 2.4 years (1.9 + 0.5)

```
       Group      V1
1: Treatment 2.40625
2:   Control 1.83250
```

```
    effect  lower_ci            p
1: 0.57375 0.2820319 0.0006955509
```

02

REPEAT EXPERIMENT 1000 TIMES

Under assumption of an **effect of 0.5 years**

```
exp.results[, mean(p<0.05)]
```

```
[1] 0.904
```

```
exp.results[, summary(effect)]
```

```
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-0.08125  0.38062  0.49875  0.50023  0.61750  1.03875
```

```
exp.results[, summary(lower_ci)]
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.3623  0.1001  0.2194  0.2197  0.3372  0.7639
```

METHODOLOGY

NO EFFECT

01

ONE TIME EXPERIMENT

Control: mean = 1.9 years
Treatment: mean = 1.9 years

```
   effect   lower_ci         p
1: -0.325 -0.6066086 0.9709938
```

02

REPEAT EXPERIMENT 1000 TIMES

Under assumption of an **no effect**

```
exp.results[, mean(p<0.05)]
```

```
[1] 0.047
```

```
exp.results[, summary(effect)]
```

```
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-0.596250 -0.116250  0.001250 -0.000505  0.113750  0.563750
```

```
exp.results[, summary(lower_ci)]
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.8615 -0.3952 -0.2802 -0.2813 -0.1652  0.2806
```

SUMMARY OF SIMULATION

Research Question	Scenario	Mean Effect in Simulated Data	95% Confidence Interval of Mean Effect	False Positive %	True Negative %	False Negative %	True Positive %
Sole Question	No Effect	-0.000505	-0.2813	4.7%	95.3%		
Sole Question	Effect of 0.5 years	0.50023	0.2197			9.6%	90.4%

SMALLER SAMPLE POPULATION?

Sample population in each group	Power
80	90%
70	86%
60	81%
50	75%
40	66%
30	55%

We simulated the expected power for smaller sample populations to pre-empt a scenario where we are unable to recruit enough employees.

We can select a power that's a good trade-off commensurate with population size.

THANK YOU

Contact us if there are any questions.