Exam 1

Denis Ostroushko

2022-10-25

Problem 1

1- A

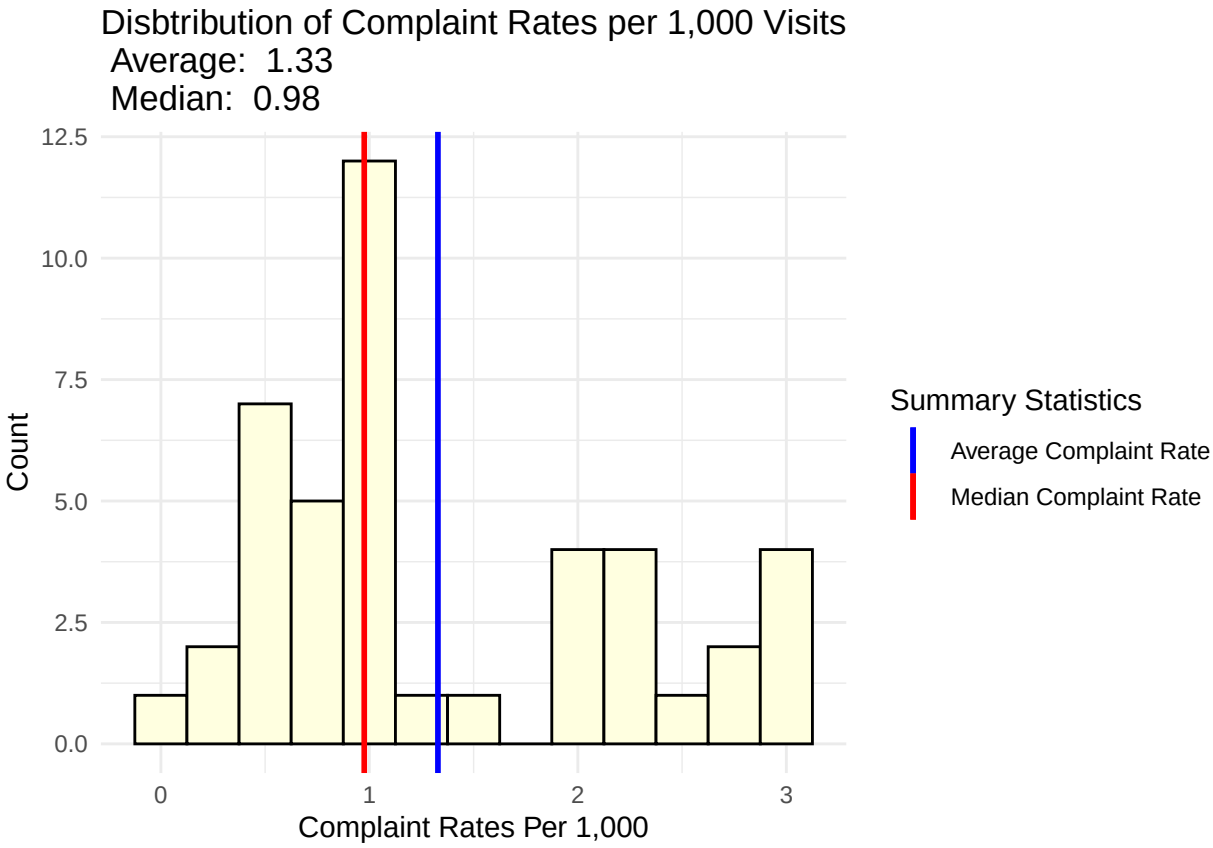
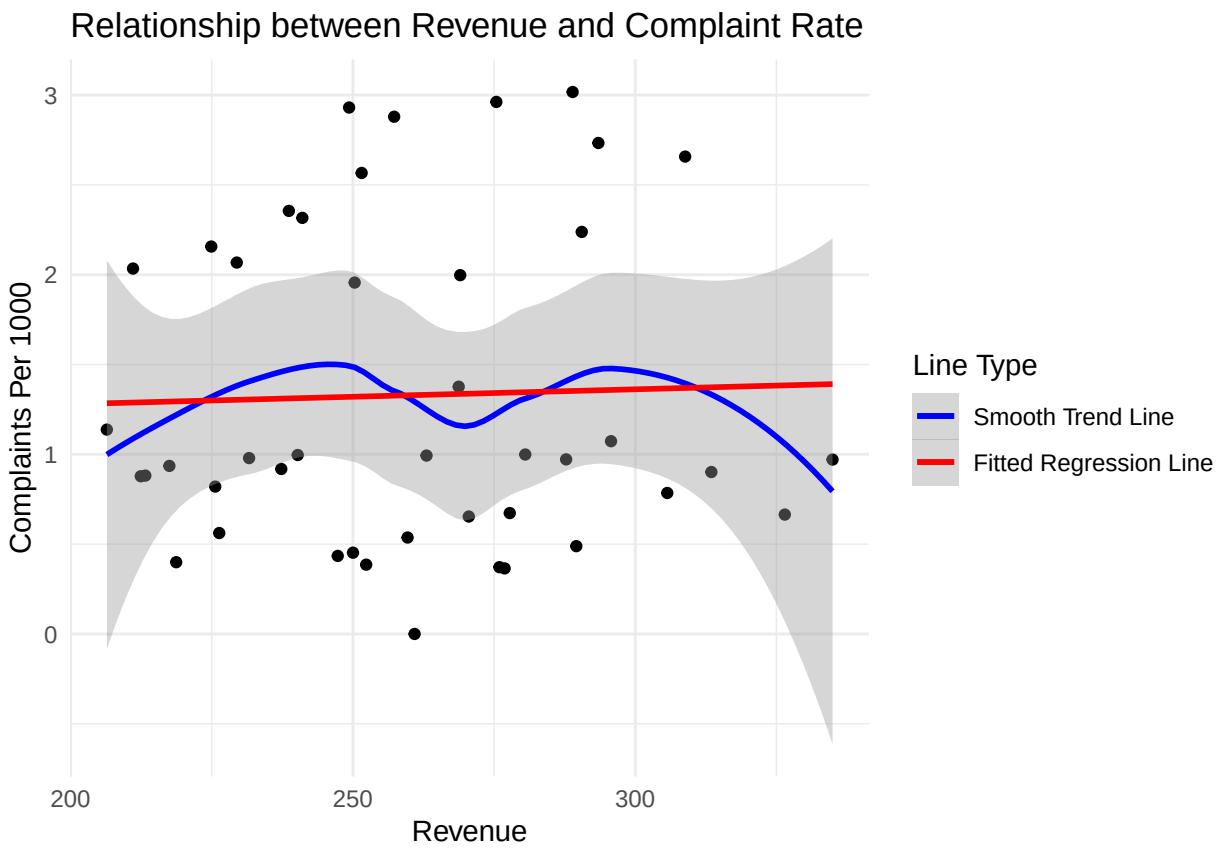


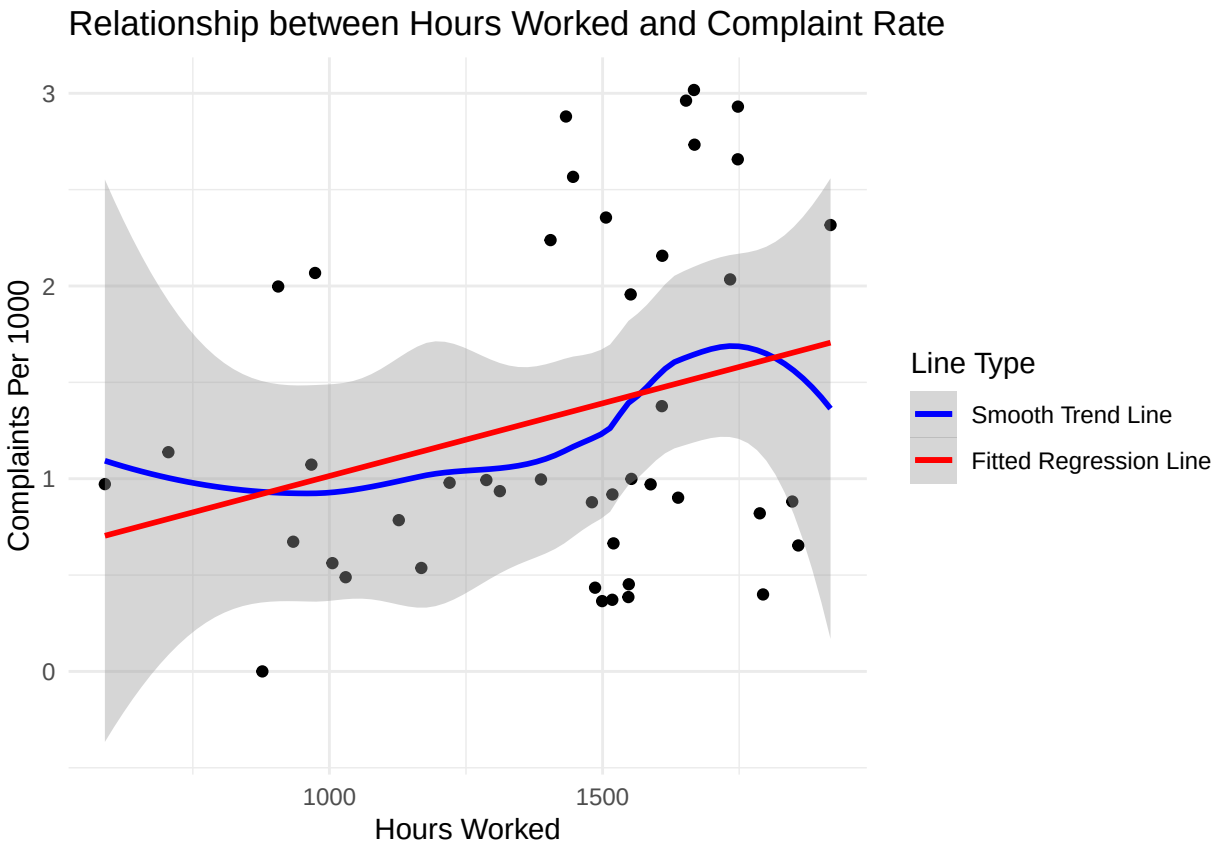
Table 1: Summary of Numeric Variables

Variables	Min	Max	Mean	S.D
complaint_rate_1000	0.00	3.02	1.33	0.88
revenue	206.42	334.94	260.14	32.64
hours	589.00	1917.25	1417.40	326.98

Table 2: Correaltion of Numeric Covariates

	Complaint Rate per 1,000	Revenue	Hours Worked
Complaint Rate per 1,000	1.0000000	0.0305876	0.2788799
Revenue	0.0305876	1.0000000	-0.0405506
Hours Worked	0.2788799	-0.0405506	1.0000000





We have categorical predictors also:

- Residency has two levels: Y, N with 54.55%, 45.45% class presence respectively
- Gender has two levels: F, M with 27.27%, 72.73% class presence respectively

Overall comments on variables

Model Assumptions

- one
- two

*three

Model Statement

$$E[Complaint\ \ Rate] = \hat{\beta}_0 + \hat{\beta}_1 * Residency + \hat{\beta}_2 * Gender + \hat{\beta}_3 * Revenue + \hat{\beta}_4 * Hours\ \ Worked$$

Overall ANOVA

Source	SSR	DF	MS	F Statistic	P(F* > F)
Regression	3.254294	4	0.8135735	1.04	0.3969
Error	30.386120	39	0.7791313	NA	NA
Total	33.640414	43	NA	NA	NA

- Null Hypothesis: $H_0 : \beta_1 = \beta_2 = ... = \beta_{p-1}$
- Alternative Hypothesis: H_a : Not all coefficients β_i are zero

- $F-$statistic: 1.04
- Cutoff F^*-statistic: 2.6123
- So, $F < F^*$, therefore we do not have enough evidence to reject the null hypothesis to conclude that some or all coefficients β_i are consistently different from zero.
- Moreover, $P(F^* > F) = 0.3969$
- Conclusion:

Regression Coefficients

Predictor	Estiamte	Standard Error	T Value	P value
(Intercept)	-0.064405	1.250366	-0.051509	0.959183
residencyY	-0.132728	0.329286	-0.403077	0.689093
genderM	0.197338	0.314907	0.626654	0.534537
revenue	0.001351	0.004610	0.293122	0.770983
hours	0.000676	0.000461	1.467079	0.150373

- R square and 0.0967
- Adjusted R Square 0.0041
- Explain Coefficients

1- B

T-test for hours worked

- Null Hypothesis: $H_0 : \hat{\beta}_4 = 0$
- Alternative Hypothesis: $H_a : \hat{\beta}_4 \neq 0$$
- Test statistic T : 1.467079
- $P(t^* > t) = 0.150373$
- Conclusion

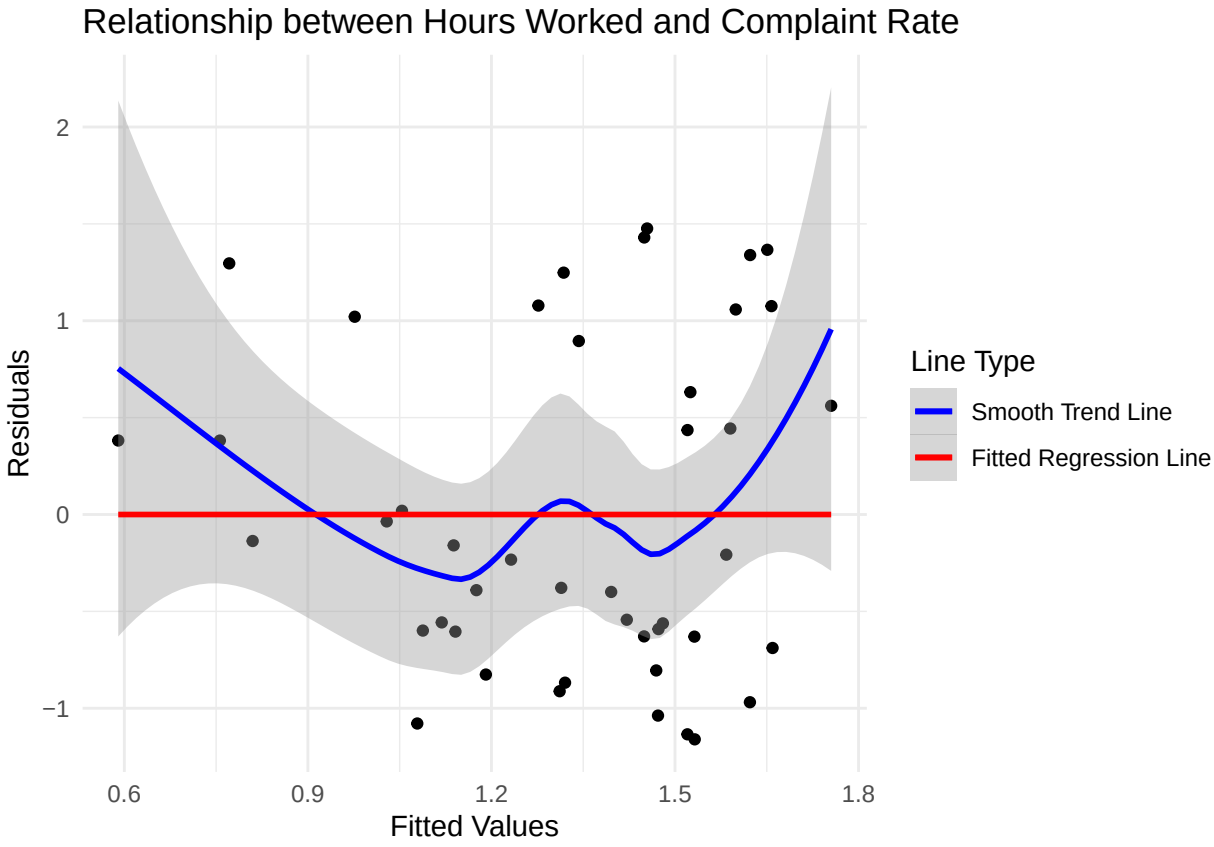
Interpretation of coefficient One additional Hour worked results in 0.000676 additional complaints on average. However, it makes more sense to say look at 100 hours, which is 0.0676

C.I.

Using formula $C.I.\ bounds = Estimate \pm 1.96 * Standard\ Error$

C.I. for the estimate 0.000676 with a 0.000461 standard error is (-0.000256, 0.001609)

1- C



1- D

Effect plots needed here, find a nice package

Problem 2

2 - A

2 - B

2 - C

2 - D