36-201 — Computer Lab #15 — Solutions

Question #2: This was an observational study, since there was no intervention controlled by the researchers (Prof. Lynch).

Question #3: YR.B1 is explanatory; it is quantitative (continuous). Is FIMR is a response variable; it is also quantitative (continuous).

Question #4: A scatter plot is a good way to explore the relationship between two continuous, quantitative variables.

♣ **Question #5:** There is a slight decrease in the overall trend (infant mortality decreases over time). There are some unusual outliers in the period from approximately 1840 to 1865.

Question #6: The "lowess" plot also shows a slight decrease; with most of the decrease happening in the period from 1850 to 1870, roughly.

Question #7: H_0 is that $\mu_{preFIMR} - \mu_{postFIMR} = 0$. H_A is that $\mu_{preFIMR} - \mu_{postFIMR} \neq 0$.

Question #8: PRE.1850 is explanatory and qualitative (discrete). FIMR is a response variable and quantitative (continuous).

Question #9: Side-by-side boxplots are a good way to explore the relationship between a discrete explanatory variable and a continuous response variable.

Question #10: The boxplots show that the median FIMR was higher before 1850 than after 1850. The range of the middle 50% of infant mortality rates after 1850 is almost completely contained between the first quartile and the median of the infant mortality rates before 1850. Both distributions look fairly symmetric. FIMR before 1850 has a moderately large high outlier; FIMR after 1850 has a mild high outlier and a mild low outlier.

Question #11: The 95% confidence interval for $\mu_{preFIMR} - \mu_{postFIMR}$ runs from 0.025 to 0.0784.

Question #12: Both ends of the confidence interval are above zero. We interpret this as saying FIMR was significantly higher before 1850 than after 1850.

♣ **Question #13:** The p-value for the test of H_0 is that $\mu_{preFIMR} - \mu_{postFIMR} = 0$, vs. H_A is that $\mu_{preFIMR} - \mu_{postFIMR} \neq 0$, is 0.0003. Small p is bad for the null hypothesis, so we reject H_0 and conclude $\mu_{preFIMR} - \mu_{postFIMR} \neq 0$ (same conclusion as with confidence interval).

In fact, $p = 0.0003$ is so much smaller than the typical cutoffs (0.10, 0.05 or 0.01) that this actually provides very strong evidence that the means are not equal.

Question #14: We conclude infant moratility did change from before 1850 to after 1850.

Question #15: The variables relevant to the second question are PRE.1850 and WORKER.

Question #16: PRE.1850 is explanatory variable; it is qualitative/discrete/categorical. WORKER is the response; it is also qualitative/discrete/categorical.

Question #17: A contingency table is an appropriate way to analyze the relationshiip between two categorical variables.

♣ **Question #18:** Percent farmers before 1850 was 47.5%, after 1850 it was 30.0%. This suggests a movement from farm work to industrial work.

Question #19: Let p_{before} and p_{after} be the true proportions of farm workers in the population before and after 1850. Then H_0 is that $p_{before} = p_{after}$; and H_A is that $p_{before} \neq p_{after}$.

Question #20: The Chi-squared statistic (sum of the squares of the standardized residuals) is 3.966. The p-value is 0.046; this is less than 0.05 so this is bad for the null hypothesis. We conclude $p_{before} \neq p_{after}$. Since the p-value is close to 0.05 we regard it as only moderately strong evidence against the null hypothesis.

Question #21: There appears to have been a a change in family occupations during the 1800s, although the evidence is only moderately strong for this.
Question #22: We have concluded that

(I) There is strong evidence that infant mortality changed during the 1800s, in particular it seems to have gone down after 1850.

(II) There is moderate evidence that the proportion of farm workers also decreased after 1850.

♣ **Question #23:** It appears that industrialization (an increase in industrial as opposed to farm work) and lower infant mortality occurred at about the same time in Swedish history, but without additional information we can't conclude that industrialization caused lower mortality rates. Lurking variables and confounding variables have not been controlled since this is an observational study and not an experiment.
Question #24: There could be a lurking variable, like increased public education, that caused both. Another possibility is that medical know-how increased at about the same time as industrialization (but not because of industrialization); this would have improved infant mortality rates as well.