

MATH 138

A Comprehensive Statistical Project (Before CH 25) Using StatCrunch

Total = 45 points

Instructions for completing Project→Answers to the project must:

1. Be submitted online, using the Assignment or Module Tools by the due date.
2. Include at least your first initial and last name at the beginning of the first page-- to avoid losing points.

BEGINNING OF PROJECT QUESTIONS/TASKS

Using StatCrunch, load the data file: CPS Wage Data From 1985. Note: The same project questions would apply even if the data were from 2005 or later, if available.

Hint: Data > Load Data > From Sample Data > Stat Crunch > **Highlight only CPS Wage Data From 1985**

- Before opening the data file, **click on “info”** and make a note of the codes for “race” and “sex”.
- After becoming familiar with the codes, click on **“close”**; click on **“Okay”** (to load the data) into StatCrunch.

Question (Q1): Classify each of the following variables as either categorical or quantitative with units **and briefly explain your classification** (to receive full credit).

[1] 1a. Education

[1] 1b. Marr

Q2: Construct and analyze the data using a contingency table:

- For the 534 persons in the CPS, construct a contingency table of **counts** with **“Race”** = row variable **and “Sex”** = column variable;
- **Hint: Stat > Table > Contingency > With data > Select row and column variables > next > Uncheck Chi-Square > calculate.**

[2] Include this table with your project submittal by (within StatCrunch), click option > copy, then paste in the document that you plan to submit. **(First of 3 attachments);**

- Use this contingency table to answers (Q)2a – (Q)2c. **Write your answer as a percent to one decimal place.**

[2] 2a. Determine the percentage of (Other and Male) from the 534 persons.

[2] 2b. Determine the percentage of female, given the Hispanics is the only Race being considered.

Problem 2 continued:

[2] 2c. [Show your work] Determine the percentage of (female or White) from the 534 persons.

Q3: Construct and analyze the data using boxplots

- For the 534 persons in the CPS, construct boxplots by:

Hint: **Graphics** > Boxplot > select Experience (work) > Group by “Sex” > check (default)-plot groups for each column > next > check both: use fences to identify outliers and draw boxes horizontally > next > create graph

Hint(This may also help): Stat > Summary Stats > Columns > Experience (Work in years) > Group by “Sex” > Accept default table groups for each column > Next > Accept the default statistics (highlighted) > calculate

[2] Include the boxplots on one pair of axes with your project submittal by (within StatCrunch), click option > copy, then paste in the document that you plan to submit. **(Second of 3 attachments);**

- Use these boxplots to answer (Q)3a – (Q)3e.

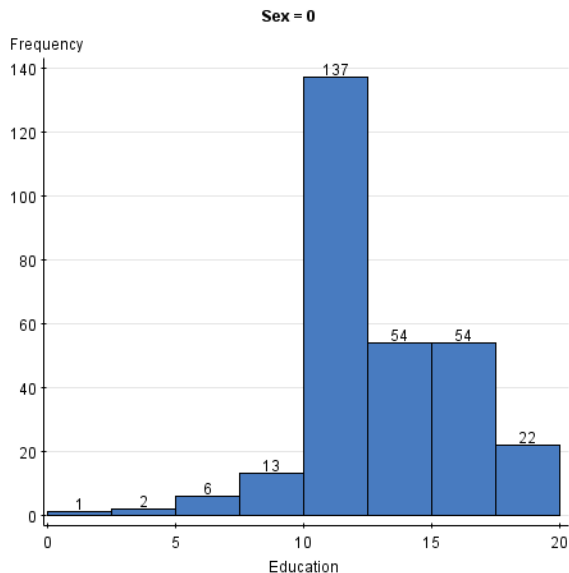
[2] 3a. Which measure of center best describes the “female” data for work experience? Explain your answer. Find the value of this statistic to one decimal place, if applicable.

[2] 3b. Which measure of spread best describes the “female” data for work experience? Explain your answer. Find the value of this statistic to one decimal place, if applicable.

[2] 3c. . Calculate the upper and lower fences for the female data, then indicate what values would be considered as outliers

[2] 3e. Which “sex” has more variation/spread? Explain your answer.

Q4: Construct and analyze the data using histograms



[2] 4a. Use the histogram above. **SHOW YOUR WORK]** What percentage of the **MALES** has more than 10 years of education? Write your final answer as a percent to one decimal place.

Q5: Applying the concepts of the Normal Distribution and the Sampling Distribution To The Data

For the 534 persons, assume the **Normal model** applies.

[2] 5a. . (Show your work by indicating the feature of the calculator used). What percent of the ages is less than 41 years (to two decimal places)? For your benefit, make a sketch. (Write your final answer as a percent to one decimal place.

Hint(This may also help): Stat > Summary Stats > Columns > Experience (Work in years) > Group by "Sex" > Accept default table groups for each column > Next > Accept the default statistics (highlighted) > calculate

[2] 5b. For this problem (5b) only, **PRETEND** that the data for all of the 534 persons are not available to you-just the data for **a sample** of 100 persons selected from the 534 persons. Assume that the mean age for this sample was 37 years, with a standard deviation of 10. What **percent** of the **mean** ages would be greater than 38 years (to two decimal places)? (Show your work by indicating the feature of the calculator used).]

Q6: Analyzing The Data Using Linear Regression

For the grouped by “Sex” variable, create a scatterplot with a fitted line plot by: **Hint:** Stat > Regression > Simple Linear > x variable = Education > y variable = Wages > Group by = Sex > next, next,... > Check “Plot Fitted Line > ..Click “next”> calculate. **Click “next” at the top of page to see the fitted line plots for males and females.** Note the codes given for “sex” on the plots. Use the linear regression to answer (Q)6a – (Q)6d).

[2] Attached this fitted line plot to your submittal by (within StatCrunch), click option > copy, then paste in the document that you plan to submit. (**Attachment 3/3**)

[2] 6a. For the “female” data, what is the value for the correlation coefficient (to two decimal places)? Does this value indicate a “strong” linear relationship between education and wages regarding females?

[2] 6b. **Assume “r” is reasonable enough**; therefore continue to answer the questions which follow: What is the linear model, **in context**, for the “females”. Write values to three decimal places.)

[2] 6c. (**Show your work**). If one the points was (14, 14.29), what is the predicted hourly wage for **females** (to two decimal places)?

[2] 6d. (**Show your work**). Determine the residual (see 6c) and indicate how well the linear model is predicting wages for “females”. (Write answer to two decimal places, to include the unit of measurement).

Q7: Analyzing The Data Using Confidence Intervals and Test of Hypotheses

For (Q)7a – (Q)7e: Assume that the data consist of a **random sample of 534 persons** from the CPS.

Problem Situation: The supervisor in charge of the CPS **hypothesized** that the **mean age** for the 534 persons would be **37.5** years of age. Test whether there is really a significant difference between the hypothesized mean age and the mean age from the sample. Use a level of significance of 0.05.

[1] 7a: What statistical test would you use for this problem situation?

Problem 7 continued:

[2] 7b. State your Null and Alternative hypotheses using English or standard statistical notation.

.

[2] 7c. State the value of the sample mean, the test statistics and the p-value for this problem (to three decimal places).

Note: p-value with at least 6 leading zeros will be considered as zero for an answer.

[2] 7d Write a conclusion regarding the null hypothesis. Be sure to include the p-value determined in 7c in your conclusion.

[2] 7e Write a 95% confidence interval (to two decimal places) using the sample mean and interpret this interval using standard statistical language. Does the 95% confidence interval help to verify your conclusion stated in 7d? Underline “Yes” or “No”, and then **explain your answer**.

END-OF PROJECT
