**Homework 5**

**SOCY 2112**

**Spring 23**

**Name:**

We have learned 3 statistical methods for testing hypotheses:

T-tests

Chi-Square

Regression

What is the correct statistical methods for testing pairs of variables at the following levels of measurement:

Variables Statistical method

1 binary variable & 1 ratio variable

2 nominal variables

2 ratio variables

1 nominal and 1 ordinal variable

1 ratio variable and 1 interval variable

1-sample test

The average educational level for men in the United States is 12.3 years. A sample of 315 men in New York state found an average educational level of 12.5 years with a standard deviation of 1.7 years. Is the NYS sample significantly different?

Is this test 1 tailed or 2 tailed?

Zobtained =

Is this significant at the 95% level?

T-tests: The following analyses are based on data from the NYC Civilian Complaint Review Board database of police officers with complaints filed against them.

Hypothesis: Female officers have fewer complaints filed against them than male officers.

|  |
| --- |
|  |
|  |  |  |
|  | *Female* | *Male* |
| Mean | 2.06724556 | 2.87322083 |
| Variance | 3.04597197 | 7.94134525 |
| Observations | 4952 | 33302 |
| Hypothesized Mean Difference | 0 |  |
| df | 9326 |  |
| t Stat | -27.586951 |  |
| P(T<=t) one-tail | 2.07E-161 |  |
| t Critical one-tail | 1.64501703 |  |
| P(T<=t) two-tail | 4.141E-161 |  |
| t Critical two-tail | 1.96021839 |   |

What are the Independent and Dependent variables?

IV

DV

Is this hypothesis 1 tailed or 2 tailed?

What is the mean number of complaints for

Males

Females

Write a short interpretation of the results, including whether the hypothesis has been proven.

Are there racial differences in the number of complaints filed against White and Latino/a officers?

|  |
| --- |
|  |
|  |  |  |
|  | *White* | *Latino/a* |
| Mean | 2.75027919 | 2.81996578 |
| Variance | 7.33420831 | 7.03310374 |
| Observations | 20595 | 9937 |
| Hypothesized Mean Difference | 0 |  |
| df | 20006 |  |
| t Stat | -2.1364947 |  |
| P(T<=t) one-tail | 0.01632561 |  |
| t Critical one-tail | 1.6449298 |  |
| P(T<=t) two-tail | 0.03265121 |  |
| t Critical two-tail | 1.96008257 |   |

What are the Independent and Dependent variables?

IV

DV

Is this hypothesis 1 tailed or 2 tailed?

What is the mean number of complaints for

White officers

Latino/a officers

Write a short interpretation of the results, including whether the hypothesis has been proven.

ChiSquare:

Cross-tabulation of gender and race among NYPD officers in the CCRB database

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Row Labels** | **Female** | **Male** | **Grand Total** |  | Expected values |  |
| Asian | 117 | 2012 | 2129 |  | 275.469805 | 1853.5302 | 2129 |
| Black | 1375 | 4176 | 5551 |  | 718.239965 | 4832.76003 | 5551 |
| Hispanic | 1863 | 8090 | 9953 |  | 1287.81163 | 8665.18837 | 9953 |
| White | 1590 | 18995 | 20585 |  | 2663.4786 | 17921.5214 | 20585 |
| **Grand Total** | **4945** | **33273** | **38218** |  | **4945** | **33273** | **38218** |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  | p value = | 0.00000000 |  |  |  |  |  |

What percent of female officers are Black?

What percent of White officers are male?

What percent of male officers are White?

What percent of Hispanic officers are female?

Looking at observed and expected values and the p value, is there a relationship between gender and race among NYPD officers in the CCRB data?

Hypothesis: There are racial differences in rank among NYPD officers in the CCRB complaints database (tables on next page)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Row Labels** | **American Indian** | **Asian** | **Black** | **Hispanic** | **White** | **Grand Total** |
| Captain |  | 44 | 46 | 78 | 312 | 480 |
| Chiefs and other ranks |  | 3 | 20 | 16 | 104 | 143 |
| Deputy Inspector |  | 13 | 17 | 30 | 133 | 193 |
| Detective | 11 | 248 | 1327 | 2303 | 4933 | 8822 |
| Inactive Ranks |  |  |  | 1 | 4 | 5 |
| Inspector |  | 3 | 19 | 14 | 115 | 151 |
| Lieutenant | 4 | 166 | 306 | 554 | 1720 | 2750 |
| Police Officer | 19 | 1259 | 2998 | 5560 | 9640 | 19476 |
| Sergeant | 2 | 395 | 820 | 1401 | 3634 | 6252 |
| **Grand Total** | **36** | **2131** | **5553** | **9957** | **20595** | **38272** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Expected values |  |  |  |  |
| 0.45150502 | 26.7265886 | 69.6446488 | 124.878763 | 258.298495 | 480 |
| 0.13451087 | 7.9622962 | 20.7483016 | 37.2034647 | 76.9514266 | 143 |
| 0.18154264 | 10.7463158 | 28.0029526 | 50.2116691 | 103.85752 | 193 |
| 8.29828595 | 491.212427 | 1280.01061 | 2295.16759 | 4747.31109 | 8822 |
| 0.00470318 | 0.27840196 | 0.72546509 | 1.30082044 | 2.69060932 | 5 |
| 0.14203595 | 8.40773934 | 21.9090458 | 39.2847774 | 81.2564015 | 151 |
| 2.58674749 | 153.121081 | 399.005801 | 715.451244 | 1479.83513 | 2750 |
| 18.3198161 | 1084.43133 | 2825.83163 | 5066.95579 | 10480.4614 | 19476 |
| 5.88085284 | 348.113817 | 907.121551 | 1626.54588 | 3364.3379 | 6252 |
| **36** | **2131** | **5553** | **9957** | **20595** | **38272** |

|  |  |
| --- | --- |
| p value = | 0.00000000000000 |

What percent of White officers are Captains?

What percent of Black officers are Police Officers?

What percent of Hispanic officers are Lieutenants?

What percent of Sergeants are Asian?

Looking at observed and expected values and the p value, is there a relationship between race and rank among NYPD officers in the CCRB data?

Correlation Matrix

The correlation matrix below is based on the NYC Community Health data (from class 5/2)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *smokes* | *binge drink* | *obesity* | *cancer screen* |  *good self-reported health* |
| smokes | 1 |  |  |  |  |
| binge drink | -0.6108603 | 1 |  |  |  |
| obesity | -0.8444837 | 0.70963196 | 1 |  |  |
| cancer screen | 0.00358152 | -0.4744067 | 0.05153097 | 1 |  |
| good health | -0.64889 | 0.85769219 | 0.64965772 | -0.5655453 | 1 |

Which two variables have the strongest correlation?

Which two variables have the weakest correlation?

What is the relationship between smoking and binge drinking?

What is the relationship between getting cancer screening and self-reported good health?

Regression hypothesis (CCRB complaint data):

officers with a high number of complaints will have a higher number of substantiated complaints. (next page)

What are the Independent and Dependent variables?

IV

DV

Is the relationship between these two variables positive or negative?

What percent of the variation in the DV comes from the IV?

Is the hypothesis proven?

Summarize the results in 1-2 sentences.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *Regression Statistics* |  |  |  |  |  |  |  |
| Multiple R | 0.54146208 |  |  |  |  |  |  |  |
| R Square | 0.29318118 |  |  |  |  |  |  |  |
| Adjusted R Square | 0.29316269 |  |  |  |  |  |  |  |
| Standard Error | 0.52641733 |  |  |  |  |  |  |  |
| Observations | 38218 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |  |  |
| Regression | 1 | 4392.72049 | 4392.72049 | 15851.604 | 0 |  |  |  |
| Residual | 38216 | 10590.2347 | 0.2771152 |  |  |  |  |  |
| Total | 38217 | 14982.9552 |   |   |   |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | -0.0833515 | 0.00384484 | -21.678809 | 0.00000000000 | -0.0908875 | -0.0758155 | -0.0908875 | -0.0758155 |
| X Variable 1 | 0.12480651 | 0.00099129 | 125.903153 | 0 | 0.12286355 | 0.12674946 | 0.12286355 | 0.12674946 |