```
-------------------------------------------------------------------------------------
> -------------------------------------------------------------------------------------
> ------------------
      name:  <unnamed>
       log:  C:\Users\saiomkark\OneDrive - The University of Chicago\AdvStats\PS7\Stat
> a_Sai_Omkar_K_PS8.log
  log type:  text
 opened on:   8 Dec 2021, 16:48:04

.
. use "C:\Users\saiomkark\OneDrive - The University of Chicago\AdvStats\PS8\homework_8
> .dta"

.
. summarize

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
     treated |        722    .4113573    .4924209          0          1
         age |        722    24.52078    6.625947         17         55
        educ |        722    10.26731    1.704774          3         16
       black |        722     .800554    .3998609          0          1
     married |        722    .1620499     .368752          0          1
-------------+---------------------------------------------------------
        hisp |        722    .1052632     .307105          0          1
        work |        722    .7285319    .4450253          0          1

.
. *1. Using the experimental data, test whether the those treated are more likely to w
> ork in the year after treatment.
. *A Use the t-test command;
.
. ttest work, by(treated) unequal level(95) welch

Two-sample t test with unequal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     425    .6964706     .022329    .4603237    .6525813    .7403599
       1 |     297    .7744108     .024294    .4186752    .7265999    .8222216
---------+--------------------------------------------------------------------
combined |     722    .7285319    .0165621    .4450253    .6960161    .7610476
---------+--------------------------------------------------------------------
    diff |           -.0779402    .0329967               -.1427288   -.0131516
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =  -2.3621
Ho: diff = 0                             Welch's degrees of freedom =  674.454

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0092         Pr(|T| > |t|) = 0.0185          Pr(T > t) = 0.9908

.
. *Observed p-value for the t-test with a 95% confidence interval is 0.0185. This valu
> e is less than 0.05. So, the Null hypothesis can be rejected and Alternative Hypothe
> sis(that the true
> difference in means of work in the year between treated and not treated people is no
> t equal to 0) cannot be rejected.
.
. *B Use the chi-square test;
. tabi 230 67 \ 296 129, chi2

           |          col
       row |         1          2 |     Total
-----------+----------------------+----------
         1 |       230         67 |       297
         2 |       296        129 |       425
-----------+----------------------+----------
     Total |       526        196 |       722

          Pearson chi2(1) =   5.3699   Pr = 0.020
```

```
.
. *Observed p-value here is 0.020 for this test. As the observed p-value is less than
> 0.05, we can reject the Null Hypothesis. And we cannot reject the alternate hypothes
> is that probabilit
> y of working in the year is same for treated and not treated is not equal.
.
. *C Use Fisher’s exact test;
. cci 230 67 296 129, exact
                                                         Proportion
                 |   Exposed   Unexposed  |      Total      exposed
-----------------+------------------------+------------------------
           Cases |       230          67  |        297       0.7744
        Controls |       296         129  |        425       0.6965
-----------------+------------------------+------------------------
           Total |       526         196  |        722       0.7285
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
      Odds ratio |         1.496067       |    1.049635    2.141541 (exact)
 Attr. frac. ex. |         .3315807       |     .047288    .5330465 (exact)
 Attr. frac. pop |         .2567797       |
                 +-------------------------------------------------
                                  1-sided Fisher's exact P = 0.0124
                                  2-sided Fisher's exact P = 0.0218

.
. *Observed p-value is 0.0218 for the two-tail fisher exact test. As this p-value is l
> ess than 0.05, we can reject the Null hypothesis that the probability of work is sam
> e for both treated
>  and not treated is equal to 1 i.w true odds ratio is equal to 1. Also, we cannot re
> ject the Alternate Hypothesis that the probability of work between people with treat
> ment and no treate
> ment is not same i.e true odds ratio is not equal to 1.
.
. *D Why do the p-values on the two-sided test differ? Which should you believe?
.
. *We observed the following p values
. * t-test : 0.0185 , Chi-square test : 0.02, Fisher exact test : 0.0218
. * As we compare the different tests, Fisher exact test is preferred as its an asympt
> otic test and fits for the binary data such as to compare the treated/non treated da
> ta. Fisher test al
> so provides an accurate significance level without relying on the assumptions where
> as asymptotic tests make assumptions about the data. When we have extremely large sa
> mple size, then it
>  is infeasable to perform Fisher test. In such situations, we can prefer Chi-square
> test as the accuracy increases as the sample size increases.
.
.
.
.
.
. *2. Using the regression command, test whether the those treated are more likely to
> work in the year after treatment.
.
. *A Regress work against the treatment indicator and test the hypothesis;
.
. regress work treated

      Source |       SS           df       MS      Number of obs   =       722
-------------+----------------------------------   F(1, 720)       =      5.40
       Model |    1.062016         1    1.062016   Prob > F        =    0.0205
    Residual |  141.730228       720  .196847539   R-squared       =    0.0074
-------------+----------------------------------   Adj R-squared   =    0.0061
       Total |  142.792244       721  .198047495   Root MSE        =    .44368

------------------------------------------------------------------------------
        work |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     treated |   .0779402   .0335553     2.32   0.020     .0120623    .1438181
       _cons |   .6964706   .0215214    32.36   0.000     .6542184    .7387228
------------------------------------------------------------------------------
```

```
.
. *Observed p-value = 0.0205 with a 95% confidence interval is less than 0.05. So, we
> can reject the null hypothesis and not the Alternate hypothesis that the true differ
> ence in means of w
> ork in year between treated and not treated is not equal to 0.
.
. *B Regress work against the treatment indicator and all other covariates in the data
>  set.
.
. regress work treated age educ black married hisp

      Source |       SS           df       MS      Number of obs   =       722
-------------+----------------------------------   F(6, 715)       =      4.71
       Model |  5.43080183         6  .905133638   Prob > F        =    0.0001
    Residual |  137.361442       715  .192113905   R-squared       =    0.0380
-------------+----------------------------------   Adj R-squared   =    0.0300
       Total |  142.792244       721  .198047495   Root MSE        =    .43831

------------------------------------------------------------------------------
        work |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     treated |   .0775439   .0332208     2.33   0.020     .0123219    .1427658
         age |  -.0033748   .0025473    -1.32   0.186    -.0083758    .0016263
        educ |   .0023976   .0097303     0.25   0.805    -.0167057    .0215009
       black |  -.1856225   .0563087    -3.30   0.001    -.2961727   -.0750724
     married |   .0526475   .0454938     1.16   0.248    -.0366699     .141965
        hisp |  -.0121452   .0742777    -0.16   0.870    -.1579737    .1336834
       _cons |   .8961161   .1315736     6.81   0.000     .6377994    1.154433
------------------------------------------------------------------------------

.
. * What happens to the R2of the regression equation?
.
. *We observed a higher adjusted R square value. This means that the additional input
> variables are adding value to the model. In univariate regression model , we observe
> d a adjused R squa
> re value of 0.0061 where are the adjusted R square value is 0.0300 with the muti var
> iate regressin model. This means a better fit of the model.
.
. *What happens to the treatment indicator? Explain why you see these results. Compare
>  them to the results in Problem 1.
.
. *Univariate model
.
. *        work |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
. *-------------+----------------------------------------------------------------
. *     treated |   .0779402   .0335553     2.32   0.020     .0120623    .1438181
.
.
. *Multivariate model
.
.
. *        work |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
. *-------------+----------------------------------------------------------------
. *     treated |   .0775439   .0332208     2.33   0.020     .0123219    .1427658
.
. *As seen above, the treatment indicator Coeff/estimate decreases slightly, but howev
> er it can still be considered that the treatment indicator still has majority of inf
> luence on the work
> .  This slight decrease may be because of the introducing other variables into the e
> quation. This is same in the case of Standard error and residual standard error wher
> e both decrease du
> e to the introduction of other variables into the regression model.
```

```
.
. * p-value t-test : 0.0185 , Chi-square test : 0.02, Fisher exact test : 0.0218
. * p-value for univariate model : 0.0205 and 0.0001 for multivariate. Comparing the p
> -values , regression is a better alternative model to the models used in the questio
> n 1.
.
.
. *3. Given the data from the experiment, do the following:
. *A Derive the marginal pmf of work when treated;
. *B Derive the marginal pmf of work when not treated;
. tab treated work, row

+----------------+
| Key            |
|----------------|
|   frequency    |
| row percentage |
+----------------+

indicator: |
      1 if |
treated, 0 |
    if not |         work
   treated |         0          1 |     Total
-----------+----------------------+----------
         0 |       129        296 |       425 
           |     30.35      69.65 |    100.00 
-----------+----------------------+----------
         1 |        67        230 |       297 
           |     22.56      77.44 |    100.00 
-----------+----------------------+----------
     Total |       196        526 |       722 
           |     27.15      72.85 |    100.00 

.
.
. *C Derive the Frechet-Hoeffding bounds for the joint distribution.
.
. * People who benefots from treatment : Bounds of joint distribution are [0.07, 0.30]
. * People who loses from the treatment : Bounds of joint distribution are [0.00, 0.23
> ]
.
.
. *4 4. Consider the following mythical experiment. There are two treatment arms: vacc
> ination with the Moderna vaccine (n = 14, 598) and vaccination with the Pfizer vacci
> ne (n = 21, 669).
> Suppose 100 of those with the Pfizer vaccine develop Covid and 269 of those with Mod
> erna vaccine have Covid.
. *A Test the hypothesis that the vaccines are equally effective against the alternati
> ve that they are not.
.
. *From above information
. *             Not affected       Affected
. *Moderna       14329              269
. *Pfizer        21569              100
.
.
. *Fisher exact test
. cci 14329 269 21569 100, exact
                                                         Proportion
                 |   Exposed   Unexposed  |      Total      exposed
-----------------+------------------------+------------------------
           Cases |     14329         269  |      14598       0.9816
        Controls |     21569         100  |      21669       0.9954
-----------------+------------------------+------------------------
           Total |     35898         369  |      36267       0.9898
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
      Odds ratio |          .246964       |    .1941703    .3121746 (exact)
 Prev. frac. ex. |          .753036       |    .6878254    .8058297 (exact)
 Prev. frac. pop |         .7495609       |
```

```
                 +---------------------------------------------------
                                  1-sided Fisher's exact P = 0.0000
                                  2-sided Fisher's exact P = 0.0000

.
. *Observed p-value 0 for two-tail fisher exact test is less than 0.05. So we can reje
> ct the null hypothesis that vaccines are equally effective (or) i.e true odds ratio
> is equal to 1)=. A
> lso we cannot reject the alternate hypothesis that the vaccines are not equally effe
> ctive (or) efficacy is not same for both i.e true odds ratio is not equal to 1.
.
.
. * B These are the actually numbers from the treated observations of the Moderna and
> Pfizer clinical trials. Explain why this is not a legitimate experimental trial.
.
. *Two arguments here. One is we donot have data of a person in two cases where in one
>  case a vaccine is administered and other case where the vaccine is not administered
> . In case, we have
>  both the data of a person, we can effectively compare the results in both the cases
>  i.e vaccinated state and unvaccinated state. As we donot have the complete data, we
>  cannot construct
> a counterfactual which deems these experimental trials cannot be considered legitima
> te. Thus, the comparision of efficacies of two vaccines cannot be considered legit.
.
. *Another argument is, because we do not have a treatment and control group. If we we
> re to have data of both the treatment and control group, then we can treat this expe
> riment triel as le
> gitimate. Thus the insights we got from comparison of efficacies of the two vaccines
>  based on only the treatment group data is not legitimate.
.
.
.
.
.
.
. log close
      name:  <unnamed>
       log:  C:\Users\saiomkark\OneDrive - The University of Chicago\AdvStats\PS7\Stat
> a_Sai_Omkar_K_PS8.log
  log type:  text
 closed on:   8 Dec 2021, 16:48:04
-------------------------------------------------------------------------------------
> -----------------------------------------------------------------------------------
> ------------------
```