

```

-----
      name: <unnamed>
      log: C:\Users\saiomkark\OneDrive - The University of Chicago\AdvStats\PS
> 6\Sai_Omkar_PS6.log
      log type: Text
      opened on: 12 Nov 2021, 23:31:14

```

```

. *
. * clear memory in stata
. *
. clear

```

```

.
. insheet using "C:\Users\saiomkark\OneDrive - The University of Chicago\AdvStat
> s\PS6\ppha312x2021.csv", names clear
(14 vars, 10,149 obs)

```

```

. label data "Data is from IPUMS-USA restricted to Albuquerque, New Mexico (2018
> )"

```

```

.
. *analyze the data using summarize to check for NAs etc..
. summarize

```

Variable	Obs	Mean	Std. Dev.	Min	Max
year	10,149	2019	0	2019	2019
statefip	0				
met2013	0				
perwt	10,149	115.5946	104.4547	2	1977
sex	0				
age	0				
race	0				
hispan	0				
bpl	0				
educd	0				
empstat	0				
uhrswork	0				
inctot	10,149	1743385	3748625	-6900	9999999
incwage	10,149	209523.8	379486.3	0	999999

```

.
. *inctot has negative values, which it cannot be. Hence keeping only those wher
> e inctot is greater than equal to 0
. keep if inctot>= 0
(9 observations deleted)

```

```

.
.
. *check the data if changes are applied
. summarize

```

Variable	Obs	Mean	Std. Dev.	Min	Max
year	10,140	2019	0	2019	2019
statefip	0				
met2013	0				
perwt	10,140	115.6266	104.4945	2	1977
sex	0				
age	0				
race	0				
hispan	0				
bpl	0				
educd	0				
empstat	0				
uhrswork	0				
inctot	10,140	1744935	3749928	0	9999999
incwage	10,140	209709.4	379603.5	0	999999

```

.
. *Encoding the string variables so that we can analyze them in STATA
. encode hispan, gen(hispan_cat)

. encode sex, gen(sex_cat)

. encode race, gen(race_cat)

. encode educd, gen(educd_cat)

. encode empstat, gen(empstat_cat)

. *encode labforce, gen(labforce_cat)
. encode age, gen(age_cat)

. encode uhrswork, gen(uhrswork_cat)

```

```

. summarize

```

Variable	Obs	Mean	Std. Dev.	Min	Max
year	10,140	2019	0	2019	2019
statefip	0				
met2013	0				
perwt	10,140	115.6266	104.4945	2	1977
sex	0				
age	0				
race	0				
hispan	0				
bpl	0				
educd	0				
empstat	0				
uhrswork	0				
inctot	10,140	1744935	3749928	0	9999999
incwage	10,140	209709.4	379603.5	0	999999
hispan_cat	10,140	2.987574	.1890676	1	5
sex_cat	10,140	1.475641	.4994309	1	2
race_cat	10,140	6.226529	3.382985	1	9
educd_cat	10,140	12.65247	9.137665	1	25
empstat_cat	10,140	1.92071	.9402849	1	4
age_cat	10,140	42.77002	24.22839	1	93
uhrswork_cat	10,140	53.46499	21.9064	1	74

```

.
.
. * variable for if respondent is Hispanic. Variable for if respondent is Africa
> n/American

```

```

. codebook hispan_cat

```

```

----- (unlabeled) -----

```

```

      type:  numeric (long)
      label:  hispan_cat

      range:  [1,5]
unique values: 5
                        units:  1
                        missing .:  0/10,140

      tabulation:  Freq.  Numeric  Label
                   4      1      Cuban
                   212     2      Mexican
                   9,851    3      Not Hispanic
                     52     4      Other
                     21     5      Puerto Rican

```

```
. gen is_hispanic = 1

. replace is_hispanic = 0 if (hispan_cat == 3)
(9,851 real changes made)
```

```
.
. codebook race_cat
```

```
-----
race_cat (unlabeled)
-----
```

```

      type:  numeric (long)
      label:  race_cat

      range:  [1,9]
unique values: 9
                        units:  1
                        missing .:  0/10,140

      tabulation:  Freq.   Numeric   Label
                   9         1   American Indian or Alaska Native
                   3,899       2   Black/African American/Negro
                   9         3   Chinese
                   3         4   Japanese
                   61        5   Other Asian or Pacific Islander
                   93        6   Other race, nec
                   15        7   Three or more major races
                   136       8   Two major races
                   5,915      9   White
```

```
. gen is_AfricanAmerican = 0

. replace is_AfricanAmerican = 1 if (race_cat == 2)
(3,899 real changes made)
```

```
.
.
. *Limit the sample to white, non-Hispanic or African American, non-Hispanic res
> pondents;
. keep if (race_cat == 9 & is_hispanic == 0) | (is_AfricanAmerican == 1 & is_his
> panic == 0)
(508 observations deleted)
```

```
.
.
. *Variable indicating whether the respondent is female
. codebook sex_cat
```

```
-----
sex_cat (unlabeled)
-----
```

```

      type:  numeric (long)
      label:  sex_cat

      range:  [1,2]
unique values: 2
                        units:  1
                        missing .:  0/9,632

      tabulation:  Freq.   Numeric   Label
                   5,054       1   Female
                   4,578       2   Male
```

```
. gen is_female = 0

. replace is_female = 1 if (sex_cat == 1)
(5,054 real changes made)

.
. * Limit the age [25:55]
. codebook age_cat
```

```
-----
age_cat (unlabeled)
-----
```

```
      type: numeric (long)
      label: age_cat, but 1 nonmissing value is not labeled

      range: [1,93]          units: 1
unique values: 93          missing.: 0/9,632

      examples: 19    26
                 37    42
                 52    56
                 65    68
```

```
. keep if (age_cat >= 25 & age_cat <= 55) & is_female == 1
(7,653 observations deleted)
```

```
.
. * Dummy variable for whether the respondent is employed.
. codebook empstat_cat
```

```
-----
empstat_cat (unlabeled)
-----
```

```
      type: numeric (long)
      label: empstat_cat

      range: [1,4]          units: 1
unique values: 4          missing.: 0/1,979

      tabulation: Freq.   Numeric   Label
                  1,407       1      Employed
                   98        2      N/A
                   414       3      Not in labor force
                   60        4      Unemployed
```

```
. gen is_employed = 0

. replace is_employed = 1 if (empstat_cat == 1)
(1,407 real changes made)
```

```
.
. *3: Using data from Homework 4, test the hypothesis that African American wome
> n and white women ages 25-55 (inclusive) have the same probability of being em
> ployed. Use a two-sided Fisher's exact test and a chi-square test with a 95% c
> onfidence level
. bysort is_AfricanAmerican: tabstat is_employed, stat(count sum)
```

```
-----
-> is_AfricanAmerican = 0
```

variable	N	sum
is_employed	1121	800

```
-----
-> is_AfricanAmerican = 1
```

variable	N	sum
is_employed	858	607

```
.
. * Null hypothesis: Hn: Probability of being employed is same for white women a
> nd African American women ages 25-55 (inclusive)
. * Alternative hypothesis: Ha: Probability of being employed is not same, great
> er for white women than that of African American women ages 25-55 (inclusive)
```

```
.
. * Fisher's test:
. cci 791 264 599 202, exact
```

	Exposed	Unexposed	Total	Proportion exposed
Cases	791	264	1055	0.7498
Controls	599	202	801	0.7478
Total	1390	466	1856	0.7489
	Point estimate		[95% Conf. Interval]	
Odds ratio	1.010409		.8125917	1.255399 (exact)
Attr. frac. ex.	.0103015		-.2306304	.2034403 (exact)
Attr. frac. pop	.0077237			
1-sided Fisher's exact P = 0.4828				
2-sided Fisher's exact P = 0.9569				

```
.
. * Observations:
. * African American women and white women ages [25:55] have the same probabilit
> y of being employed. The p-value = 0.9569 for two-tail test and The p-value =
> 0.4828 for the one-tail test are greater than 0.05. Thus the Null Hypothesis
> is not rejected.
```

```
.
. * chi-square test:
. tabi 791 264 \ 599 202, chi2
```

row	col 1	col 2	Total
1	791	264	1,055
2	599	202	801
Total	1,390	466	1,856
Pearson chi2(1) = 0.0092 Pr = 0.924			

```
.
. * Observations:
. * African American women and white women ages [25:55] have the same probability
> of being employed. The p-value = 0.924 for the test is greater than 0.05. Thu
> s the Null Hypothesis cannot be rejected.
```

```
.
. * Question 4: Using data from Homework 4, test the hypothesis that African Ame
> rican women and white women ages 25-55 (inclusive) have the same total income.
> Use a two-sided t-test with a 95% confidence level. How confident of this tes
> t are you? Why?
```

```

. gen inc_tot_employed = inctot * is_employed*is_AfricanAmerican*is_female
. ttest inc_tot_employed, by(is_AfricanAmerican) unequal level(95) welch

Two-sample t test with unequal variances
-----
      Group |      Obs      Mean      Std. Err.      Std. Dev.      [95% Conf. Interval]
-----+-----
          0 |    1,121          0          0          0          0          0
          1 |      858    29232.58    1132.77    33180.69    27009.25    31455.91
-----+-----
combined |    1,979    12673.85    589.1802    26210.24    11518.37    13829.33
-----+-----
      diff |      -29232.58    1132.77          -31455.91    -27009.25
-----+-----
      diff = mean(0) - mean(1)                                t = -25.8063
Ho: diff = 0                                           Welch's degrees of freedom =      857

      Ha: diff < 0                                Ha: diff != 0                                Ha: diff > 0
Pr(T < t) = 0.0000                                Pr(|T| > |t|) = 0.0000                                Pr(T > t) = 1.0000

.
. * Observations for two-sided t-test:
. * P-value = 0 for two-sided t-test with a 95% confidence level is less than 0
> .05. Thus the Null hypothesis that true difference in means of total income is
> equal to 0 between white women and African/American women ages [25:55] can be
> rejected. The Alternative Hypothesis that true difference in means of total i
> ncome is not 0 between white women and African American women ages [25:55] can
> not be rejected.
.
. *Similarly, for difference in mean (White - African/American) is "less" than z
> ero. We got P =1 with 95% confidence level which is greater than 0.05. Thus we
> can reject the alternate hypotheses that the true difference in mean of total
> income between White women and African/American women with age range [25:55]
> is less than zero can be rejected. Null hypotheses that true difference in the
> above said mean is zero cannot be rejected.
.
. *Similarly, the alternate hypotheses that difference in mean of total income f
> or White women and African/American women with age range [25:55] is greater th
> an 0 cannot be rejected because of the p-value = 0 for 95% confidence level ,
> which is less than 0.05. Also, due to the p value we can reject the null hypot
> heses that the true difference of the above said mean can be rejected.
.
.
. * Looking at the above p values in different observances, the hypotheses that
> the the mean of total income of White women is greater than the African/Americ
> an women with age ranges [25-55] cannot be rejected.
.
.
. log close
      name: <unnamed>
      log: C:\Users\saiomkark\OneDrive - The University of Chicago\AdvStats\PS
> 6\Sai_Omkar_PS6.log
      log type: text
      closed on: 12 Nov 2021, 23:31:14
-----

```