Part 1: 1. tabulate ridreth1

race/hispan |

ic origin | Freq. Percent Cum.

------------+-----------------------------------

1 | 425 9.91 9.91

2 | 407 9.49 19.39

3 | 1,714 39.95 59.35

4 | 1,095 25.52 84.87

5 | 649 15.13 100.00

------------+-----------------------------------

Total | 4,290 100.00

tabulate ridreth1 dmdeduc2

race/hispa | education level - adults 20+

nic origin | 1 2 3 4 5 | Total

-----------+-------------------------------------------------------+----------

1 | 123 86 89 93 34 | 425

2 | 82 71 83 106 65 | 407

3 | 68 197 363 577 509 | 1,714

4 | 57 168 277 382 211 | 1,095

5 | 31 42 86 171 319 | 649

-----------+-------------------------------------------------------+----------

Total | 361 564 898 1,329 1,138 | 4,290

. tabulate ridreth1 riagendr

race/hispa | gender

nic origin | 1 2 | Total

-----------+----------------------+----------

1 | 228 197 | 425

2 | 193 214 | 407

3 | 867 847 | 1,714

4 | 507 588 | 1,095

5 | 337 312 | 649

-----------+----------------------+----------

Total | 2,132 2,158 | 4,290

tabulate ridreth1 ridageyr

race/hispa | age in years at screening

nic origin | 20 21 22 23 24 | Total

-----------+-------------------------------------------------------+----------

1 | 8 15 9 4 7 | 425

2 | 6 10 8 9 6 | 407

3 | 30 16 27 27 28 | 1,714

4 | 30 28 24 20 28 | 1,095

5 | 11 17 16 17 30 | 649

-----------+-------------------------------------------------------+----------

Total | 85 86 84 77 99 | 4,290

race/hispa | age in years at screening

nic origin | 25 26 27 28 29 | Total

-----------+-------------------------------------------------------+----------

1 | 8 8 15 7 10 | 425

2 | 4 7 5 9 7 | 407

3 | 28 18 32 20 25 | 1,714

4 | 23 9 13 14 24 | 1,095

5 | 14 13 11 16 18 | 649

-----------+-------------------------------------------------------+----------

Total | 77 55 76 66 84 | 4,290

race/hispa | age in years at screening

nic origin | 30 31 32 33 34 | Total

-----------+-------------------------------------------------------+----------

1 | 6 9 11 9 9 | 425

2 | 8 13 4 8 5 | 407

3 | 37 42 32 26 32 | 1,714

4 | 17 17 17 16 12 | 1,095

5 | 16 13 10 13 13 | 649

-----------+-------------------------------------------------------+----------

Total | 84 94 74 72 71 | 4,290

race/hispa | age in years at screening

nic origin | 35 36 37 38 39 | Total

-----------+-------------------------------------------------------+----------

1 | 5 18 15 11 7 | 425

2 | 3 4 4 6 8 | 407

3 | 36 30 32 20 32 | 1,714

4 | 14 14 14 12 9 | 1,095

5 | 11 23 11 12 20 | 649

-----------+-------------------------------------------------------+----------

Total | 69 89 76 61 76 | 4,290

race/hispa | age in years at screening

nic origin | 40 41 42 43 44 | Total

-----------+-------------------------------------------------------+----------

1 | 11 9 13 9 5 | 425

2 | 12 6 4 3 8 | 407

3 | 30 21 25 29 27 | 1,714

4 | 16 17 19 16 19 | 1,095

5 | 15 14 3 15 6 | 649

-----------+-------------------------------------------------------+----------

Total | 84 67 64 72 65 | 4,290

race/hispa | age in years at screening

nic origin | 45 46 47 48 49 | Total

-----------+-------------------------------------------------------+----------

1 | 4 6 7 7 8 | 425

2 | 6 6 4 4 3 | 407

3 | 26 34 27 20 36 | 1,714

4 | 19 15 24 19 17 | 1,095

5 | 13 12 18 12 10 | 649

-----------+-------------------------------------------------------+----------

Total | 68 73 80 62 74 | 4,290

race/hispa | age in years at screening

nic origin | 50 51 52 53 54 | Total

-----------+-------------------------------------------------------+----------

1 | 9 7 7 7 14 | 425

2 | 14 12 6 2 7 | 407

3 | 33 24 25 24 23 | 1,714

4 | 22 25 26 22 23 | 1,095

5 | 6 11 13 9 14 | 649

-----------+-------------------------------------------------------+----------

Total | 84 79 77 64 81 | 4,290

race/hispa | age in years at screening

nic origin | 55 56 57 58 59 | Total

-----------+-------------------------------------------------------+----------

1 | 3 6 5 8 4 | 425

2 | 7 9 5 4 8 | 407

3 | 28 26 26 19 23 | 1,714

4 | 18 25 14 19 17 | 1,095

5 | 13 11 5 10 6 | 649

-----------+-------------------------------------------------------+----------

Total | 69 77 55 60 58 | 4,290

race/hispa | age in years at screening

nic origin | 60 61 62 63 64 | Total

-----------+-------------------------------------------------------+----------

1 | 11 6 11 8 7 | 425

2 | 19 11 7 10 10 | 407

3 | 29 18 23 35 29 | 1,714

4 | 27 33 29 30 22 | 1,095

5 | 7 9 14 8 7 | 649

-----------+-------------------------------------------------------+----------

Total | 93 77 84 91 75 | 4,290

race/hispa | age in years at screening

nic origin | 65 66 67 68 69 | Total

-----------+-------------------------------------------------------+----------

1 | 6 6 3 4 1 | 425

2 | 10 6 9 7 7 | 407

3 | 29 10 12 22 18 | 1,714

4 | 20 18 20 18 13 | 1,095

5 | 8 8 4 4 8 | 649

-----------+-------------------------------------------------------+----------

Total | 73 48 48 55 47 | 4,290

race/hispa | age in years at screening

nic origin | 70 71 72 73 74 | Total

-----------+-------------------------------------------------------+----------

1 | 2 1 2 0 4 | 425

2 | 9 3 4 5 3 | 407

3 | 33 23 24 22 29 | 1,714

4 | 15 8 7 9 8 | 1,095

5 | 4 3 3 3 3 | 649

-----------+-------------------------------------------------------+----------

Total | 63 38 40 39 47 | 4,290

race/hispa | age in years at screening

nic origin | 75 76 77 78 79 | Total

-----------+-------------------------------------------------------+----------

1 | 3 0 1 2 0 | 425

2 | 4 1 1 3 2 | 407

3 | 20 18 14 15 18 | 1,714

4 | 8 7 8 6 13 | 1,095

5 | 2 3 5 5 2 | 649

-----------+-------------------------------------------------------+----------

Total | 37 29 29 31 35 | 4,290

| age in

| years at

race/hispa | screening

nic origin | 80 | Total

-----------+-----------+----------

1 | 7 | 425

2 | 12 | 407

3 | 177 | 1,714

4 | 29 | 1,095

5 | 18 | 649

-----------+-----------+----------

Total | 243 | 4,290

tabulate ridreth1 dbq700

race/hispa | how healthy is the diet

nic origin | 1 2 3 4 5 | Total

-----------+-------------------------------------------------------+----------

1 | 26 44 167 157 31 | 425

2 | 29 68 200 91 19 | 407

3 | 172 424 732 317 69 | 1,714

4 | 91 185 482 254 83 | 1,095

5 | 90 197 270 70 22 | 649

-----------+-------------------------------------------------------+----------

Total | 408 918 1,851 889 224 | 4,290

tabulate ridreth1 dsdcount

race/hispa | total # of dietary supplements taken

nic origin | 0 1 2 3 4 | Total

-----------+-------------------------------------------------------+----------

1 | 284 75 32 17 8 | 425

2 | 235 75 42 27 15 | 407

3 | 729 400 232 135 75 | 1,714

4 | 610 253 108 61 20 | 1,095

5 | 302 165 80 41 25 | 649

-----------+-------------------------------------------------------+----------

Total | 2,160 968 494 281 143 | 4,290

race/hispa | total # of dietary supplements taken

nic origin | 5 6 7 8 9 | Total

-----------+-------------------------------------------------------+----------

1 | 3 4 1 0 1 | 425

2 | 6 3 3 1 0 | 407

3 | 49 34 19 14 9 | 1,714

4 | 20 9 3 3 2 | 1,095

5 | 17 4 5 4 4 | 649

-----------+-------------------------------------------------------+----------

Total | 95 54 31 22 16 | 4,290

race/hispa | total # of dietary supplements taken

nic origin | 10 11 12 13 14 | Total

-----------+-------------------------------------------------------+----------

1 | 0 0 0 0 0 | 425

2 | 0 0 0 0 0 | 407

3 | 6 4 3 0 1 | 1,714

4 | 4 1 0 0 0 | 1,095

5 | 0 0 0 2 0 | 649

-----------+-------------------------------------------------------+----------

Total | 10 5 3 2 1 | 4,290

| total # of dietary supplements

race/hispa | taken

nic origin | 15 16 18 | Total

-----------+---------------------------------+----------

1 | 0 0 0 | 425

2 | 0 0 0 | 407

3 | 1 1 2 | 1,714

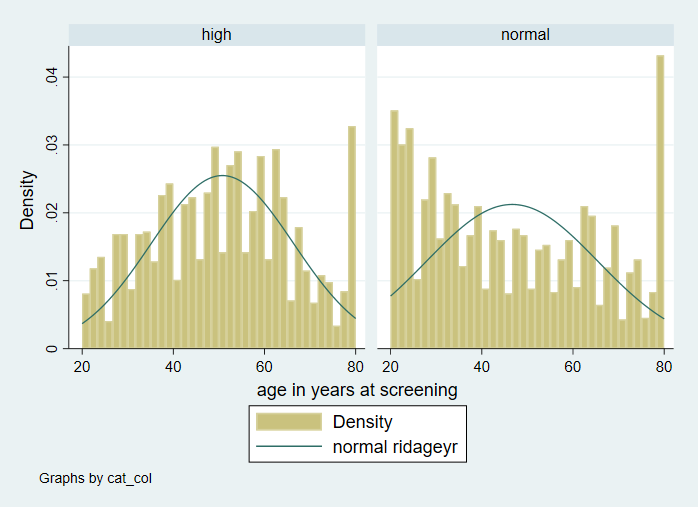
4 | 1 0 0 | 1,095

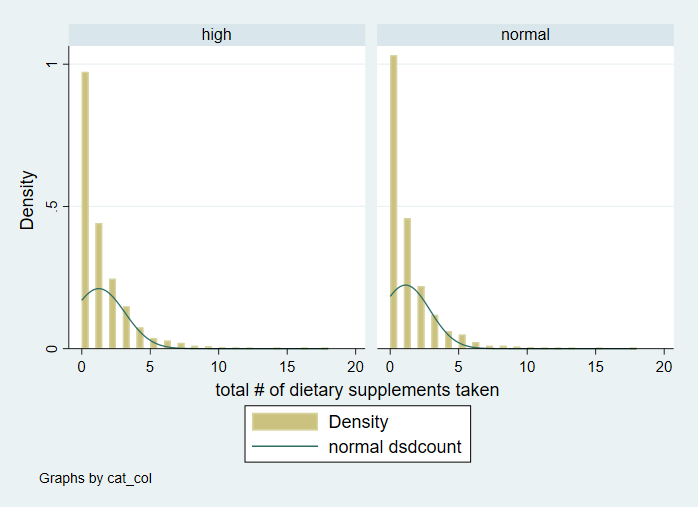
5 | 0 0 0 | 649

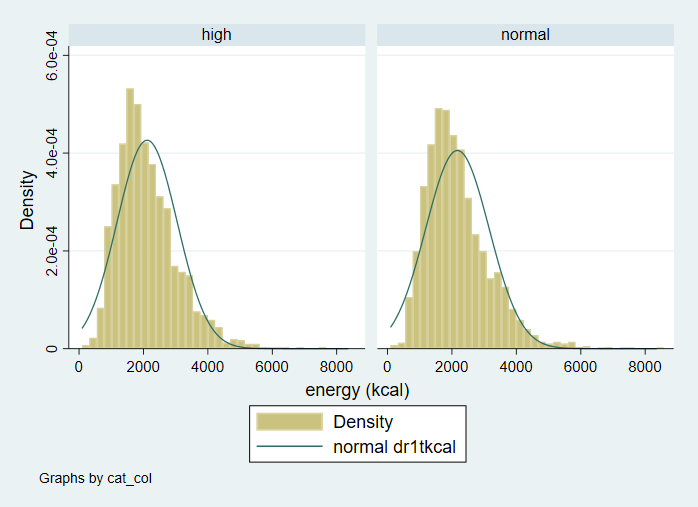
-----------+---------------------------------+----------

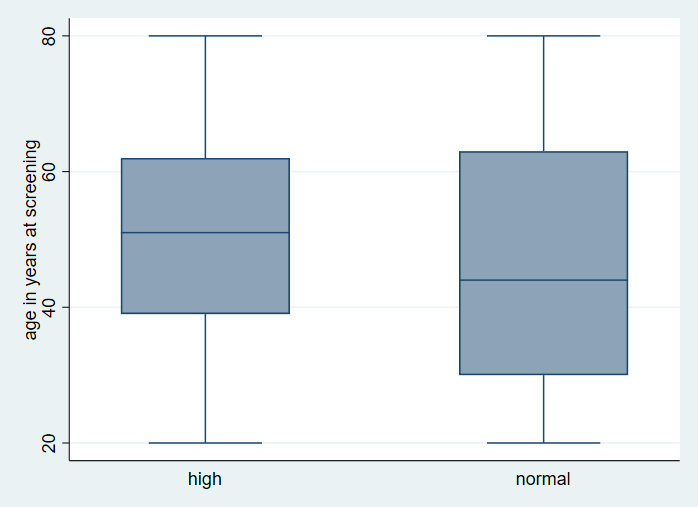
Total | 2 1 2 | 4,290

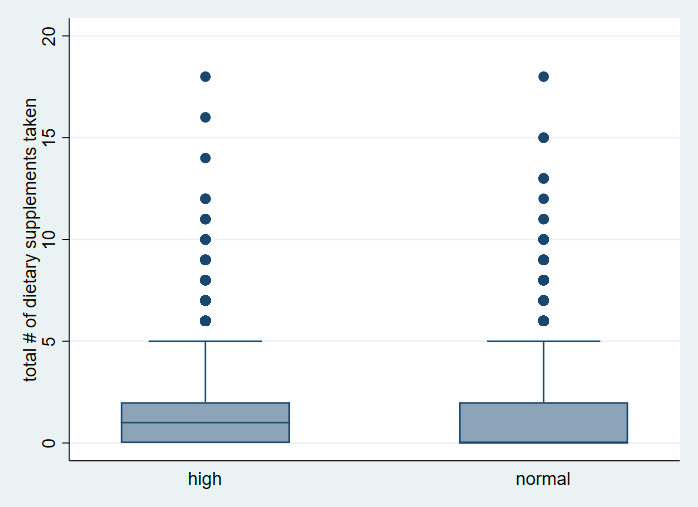
.

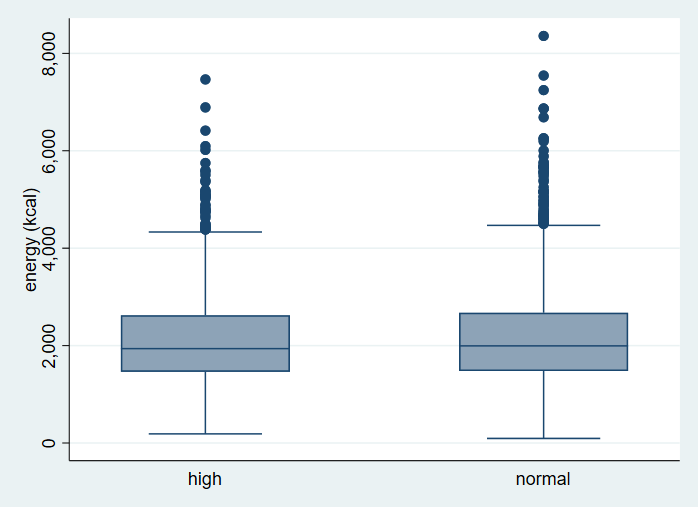


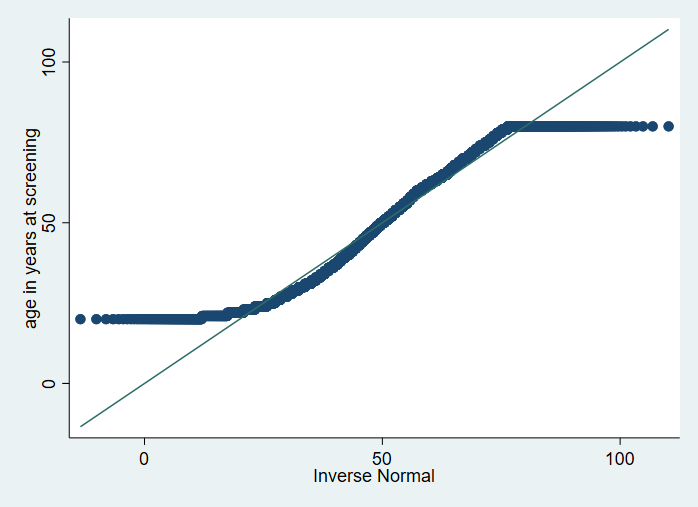


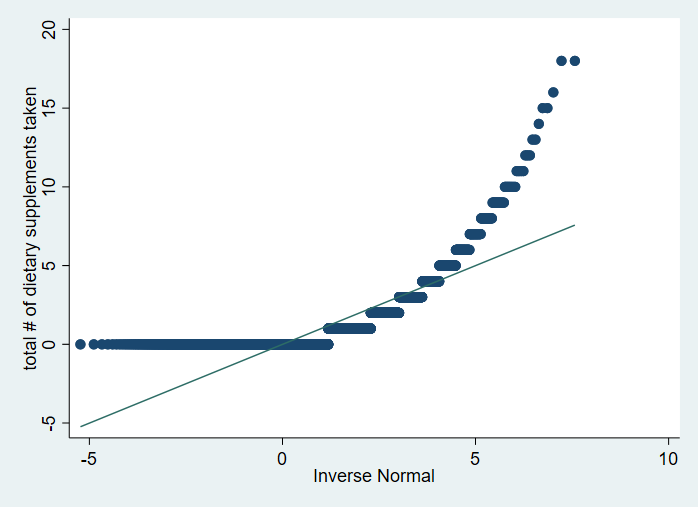


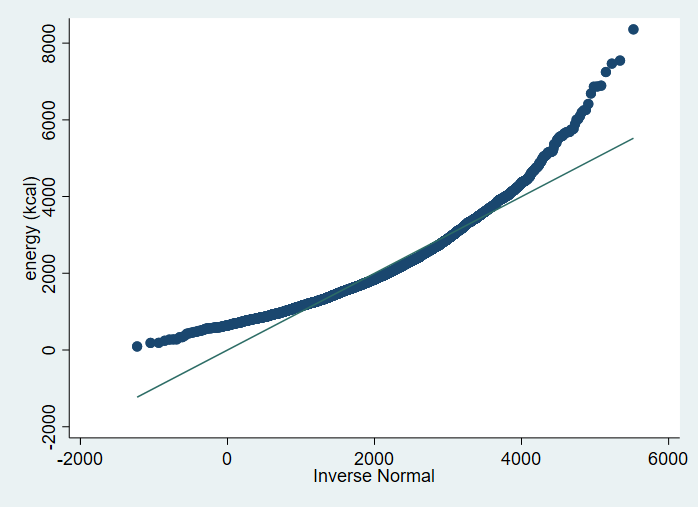












b. summarize ridageyr, detail

age in years at screening

-------------------------------------------------------------

Percentiles Smallest

1% 20 20

5% 22 20

10% 24 20 Obs 4,290

25% 33 20 Sum of wgt. 4,290

50% 48 Mean 48.35641

Largest Std. dev. 17.66649

75% 62 80

90% 74 80 Variance 312.1049

95% 80 80 Skewness .1542648

99% 80 80 Kurtosis 1.912001

summarize dsdcount, detail

total # of dietary supplements taken

-------------------------------------------------------------

Percentiles Smallest

1% 0 0

5% 0 0

10% 0 0 Obs 4,290

25% 0 0 Sum of wgt. 4,290

50% 0 Mean 1.170163

Largest Std. dev. 1.829385

75% 2 15

90% 3 16 Variance 3.34665

95% 5 18 Skewness 2.779734

99% 8 18 Kurtosis 14.72598

The median for the variable "total # of dietary supplements taken" is 0, and the interquartile range (IQR) is the difference between the 75th percentile (2) and the 25th percentile (0), which is 2. Therefore, the appropriate measure of location is the median, and the appropriate measure of dispersion is the IQR.

summarize dr1tkcal, detail

energy (kcal)

-------------------------------------------------------------

Percentiles Smallest

1% 599 93

5% 892 186

10% 1092.5 188 Obs 4,290

25% 1469 242 Sum of wgt. 4,290

50% 1972.5 Mean 2146.335

Largest Std. dev. 964.5429

75% 2646 7247

90% 3437.5 7466 Variance 930343

95% 3947 7546 Skewness 1.141869

99% 5164 8359 Kurtosis 5.256997

For age in years at screening, the mean is 48.36 years with a standard deviation of 17.67 years. The median is 48 years and the interquartile range (IQR) is 29 years. The distribution appears to be slightly positively skewed but relatively close to a normal distribution.

For total number of dietary supplements taken, the median is 0 with an IQR of 2 supplements. The mean is 1.17 supplements with a standard deviation of 1.83 supplements. The distribution appears to be highly skewed to the right, with a long tail of high values.

For energy intake in kcal, the mean is 2146.34 kcal with a standard deviation of 964.54 kcal. The median is 1972.5 kcal and the IQR is 1176.5 kcal. The distribution appears to be slightly positively skewed but relatively close to a normal distribution.

Since age and energy intake appear to be approximately normally distributed, we can use the mean and standard deviation as measures of location and dispersion, respectively. However, since the distribution of total number of dietary supplements taken appears to be highly skewed, we should use the median and IQR instead.

For testing the simple association between each independent variable and high versus normal total cholesterol subjects, we can use the t-test for age and energy intake since they are approximately normally distributed, and the Wilcoxon rank-sum test for total number of dietary supplements taken since it is highly skewed.

c. For the variable "age in years at screening", the data appear to be roughly symmetric with a small positive skewness and a kurtosis close to 3, which is the kurtosis value for a normal distribution. Therefore, a parametric test such as a t-test or ANOVA could be used for this variable.

For the variable "total # of dietary supplements taken", the data have a large number of values at 0 and a right-skewed distribution, with a high kurtosis value. This suggests that the data may not follow a normal distribution, and a nonparametric test such as the Wilcoxon rank-sum test or Kruskal-Wallis test may be more appropriate.

For the variable "energy (kcal)", the data appear to be moderately right-skewed with a kurtosis value greater than 3, which may suggest that the data do not follow a normal distribution. However, with a large sample size of 4,290, the Central Limit Theorem may apply and a parametric test such as a t-test or ANOVA could still be used.

c. ttest ridageyr, by(cat\_col)

Two-sample t test with equal variances

------------------------------------------------------------------------------

Group | Obs Mean Std. err. Std. dev. [95% conf. interval]

---------+--------------------------------------------------------------------

high | 1,776 50.76014 .3712728 15.64641 50.03196 51.48831

normal | 2,514 46.65831 .3746084 18.78279 45.92374 47.39289

---------+--------------------------------------------------------------------

Combined | 4,290 48.35641 .2697253 17.66649 47.82761 48.88521

---------+--------------------------------------------------------------------

diff | 4.101822 .5440843 3.035135 5.168508

------------------------------------------------------------------------------

diff = mean(high) - mean(normal) t = 7.5389

H0: diff = 0 Degrees of freedom = 4288

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0

Pr(T < t) = 1.0000 Pr(|T| > |t|) = 0.0000 Pr(T > t) = 0.0000

ttest dr1tkcal, by(cat\_col)

Two-sample t test with equal variances

------------------------------------------------------------------------------

Group | Obs Mean Std. err. Std. dev. [95% conf. interval]

---------+--------------------------------------------------------------------

high | 1,776 2113.055 22.19807 935.4848 2069.518 2156.592

normal | 2,514 2169.845 19.62633 984.0603 2131.359 2208.33

---------+--------------------------------------------------------------------

Combined | 4,290 2146.335 14.72627 964.5429 2117.464 2175.206

---------+--------------------------------------------------------------------

diff | -56.78969 29.88918 -115.3879 1.808565

------------------------------------------------------------------------------

diff = mean(high) - mean(normal) t = -1.9000

H0: diff = 0 Degrees of freedom = 4288

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0

Pr(T < t) = 0.0287 Pr(|T| > |t|) = 0.0575 Pr(T > t) = 0.9713

ranksum dsdcount, by(cat\_col)

Two-sample Wilcoxon rank-sum (Mann–Whitney) test

cat\_col | Obs Rank sum Expected

-------------+---------------------------------

high | 1776 3899500 3810408

normal | 2514 5304695 5393787

-------------+---------------------------------

Combined | 4290 9204195 9204195

Unadjusted variance 1.597e+09

Adjustment for ties -2.251e+08

----------

Adjusted variance 1.371e+09

H0: dsdcount(cat\_col==high) = dsdcount(cat\_col==normal)

z = 2.406

Prob > |z| = 0.0161

tab riagendr

gender | Freq. Percent Cum.

------------+-----------------------------------

1 | 2,132 49.70 49.70

2 | 2,158 50.30 100.00

------------+-----------------------------------

Total | 4,290 100.00

.

. tab ridreth1

race/hispan |

ic origin | Freq. Percent Cum.

------------+-----------------------------------

1 | 425 9.91 9.91

2 | 407 9.49 19.39

3 | 1,714 39.95 59.35

4 | 1,095 25.52 84.87

5 | 649 15.13 100.00

------------+-----------------------------------

Total | 4,290 100.00

.

. tab dmdeduc2

education |

level - |

adults 20+ | Freq. Percent Cum.

------------+-----------------------------------

1 | 361 8.41 8.41

2 | 564 13.15 21.56

3 | 898 20.93 42.49

4 | 1,329 30.98 73.47

5 | 1,138 26.53 100.00

------------+-----------------------------------

Total | 4,290 100.00

tabulate riagendr cat\_col, chi2

| cat\_col

gender | high normal | Total

-----------+----------------------+----------

1 | 798 1,334 | 2,132

2 | 978 1,180 | 2,158

-----------+----------------------+----------

Total | 1,776 2,514 | 4,290

Pearson chi2(1) = 27.5203 Pr = 0.000

.

. tabulate ridreth1 cat\_col, chi2

race/hispa | cat\_col

nic origin | high normal | Total

-----------+----------------------+----------

1 | 191 234 | 425

2 | 190 217 | 407

3 | 727 987 | 1,714

4 | 420 675 | 1,095

5 | 248 401 | 649

-----------+----------------------+----------

Total | 1,776 2,514 | 4,290

Pearson chi2(4) = 14.5072 Pr = 0.006

.

. tabulate dmdeduc2 cat\_col, chi2

education |

level - | cat\_col

adults 20+ | high normal | Total

-----------+----------------------+----------

1 | 149 212 | 361

2 | 238 326 | 564

3 | 373 525 | 898

4 | 548 781 | 1,329

5 | 468 670 | 1,138

-----------+----------------------+----------

Total | 1,776 2,514 | 4,290

Pearson chi2(4) = 0.2082 Pr = 0.995

.

tabulate riagendr cat\_col, exact

| cat\_col

gender | high normal | Total

-----------+----------------------+----------

1 | 798 1,334 | 2,132

2 | 978 1,180 | 2,158

-----------+----------------------+----------

Total | 1,776 2,514 | 4,290

Fisher's exact = 0.000

1-sided Fisher's exact = 0.000

.

. tabulate ridreth1 cat\_col, exact

Enumerating sample-space combinations:

stage 5: enumerations = 1

stage 4: enumerations = 72

stage 3: enumerations = 4120

stage 2: enumerations = 236595

stage 1: enumerations = 0

race/hispa | cat\_col

nic origin | high normal | Total

-----------+----------------------+----------

1 | 191 234 | 425

2 | 190 217 | 407

3 | 727 987 | 1,714

4 | 420 675 | 1,095

5 | 248 401 | 649

-----------+----------------------+----------

Total | 1,776 2,514 | 4,290

Fisher's exact = 0.006

.

. tabulate dmdeduc2 cat\_col, exact

Enumerating sample-space combinations:

stage 5: enumerations = 1

stage 4: enumerations = 8

stage 3: enumerations = 63

stage 2: enumerations = 498

stage 1: enumerations = 0

education |

level - | cat\_col

adults 20+ | high normal | Total

-----------+----------------------+----------

1 | 149 212 | 361

2 | 238 326 | 564

3 | 373 525 | 898

4 | 548 781 | 1,329

5 | 468 670 | 1,138

-----------+----------------------+----------

Total | 1,776 2,514 | 4,290

Fisher's exact = 0.995

PART 2

tabstat dr1tsugr chol\_cat, stat(mean): This output shows the mean (average) value of the variable "dr1tsugr" for each category of the variable "chol\_cat".

The first row of the table indicates that the mean value of "dr1tsugr" for subjects with normal total cholesterol levels is 116.5496, while the second row indicates that the mean value of "dr1tsugr" for subjects with high total cholesterol levels is 1.470862.

This suggests that there may be a difference in the mean sugar intake between those with normal and high total cholesterol levels, with the latter group having a lower mean sugar intake. However, further statistical analysis would be needed to confirm whether this difference is statistically significant.

Stats | dr1tsugr chol\_cat

---------+--------------------

Mean | 116.5496 1.470862

tabstat dr1ttfat chol\_cat, stat(mean): The output you provided shows the mean (average) value of the variable "dr1ttfat" (total fat intake in grams per day) for each category of the variable "chol\_cat" (the new categorical variable that groups individuals by their total cholesterol levels).

The mean value of "dr1ttfat" for all individuals in the dataset is 79.73834 grams per day. The mean value of "dr1ttfat" for individuals in the normal cholesterol category (chol\_cat=1) and high cholesterol category (chol\_cat=3) is slightly higher than the overall mean, indicating that individuals in these categories tend to consume slightly more total fat per day. The mean value of "dr1ttfat" for individuals in the borderline cholesterol category (chol\_cat=2) is slightly lower than the overall mean, indicating that individuals in this category tend to consume slightly less total fat per day. However, it is important to note that these differences may or may not be statistically significant and further analysis would be needed to determine this.

Stats | dr1ttfat chol\_cat

---------+--------------------

Mean | 79.73834 1.470862

b. The statistical hypothesis for question (i) is:

H0: The mean total sugar intake is the same among the three cholesterol levels.

Ha: The mean total sugar intake is different among the three cholesterol levels.

The appropriate hypothesis test to use is one-way analysis of variance (ANOVA) because we are comparing means of a continuous variable (total sugar intake) across three groups (cholesterol levels). The assumptions for ANOVA include normality, homogeneity of variances, and independence. These assumptions can be checked using normal probability plots, Levene's test for homogeneity of variances, and residual plots.

The statistical hypothesis for question (ii) is:

H0: The mean total fat intake is the same among the three cholesterol levels.

Ha: The mean total fat intake is different among the three cholesterol levels.

The appropriate hypothesis test to use is also one-way ANOVA because we are comparing means of a continuous variable (total fat intake) across three groups (cholesterol levels). The assumptions for ANOVA include normality, homogeneity of variances, and independence, which can be checked as described above.

c. anova dr1tsugr chol\_cat

Number of obs = 4,290 R-squared = 0.0011

Root MSE = 77.2144 Adj R-squared = 0.0006

Source | Partial SS df MS F Prob>F

-----------+----------------------------------------------------

Model | 27954.115 2 13977.057 2.34 0.0960

|

chol\_cat | 27954.115 2 13977.057 2.34 0.0960

|

Residual | 25559369 4,287 5962.0642

-----------+----------------------------------------------------

Total | 25587323 4,289 5965.8017

oneway dr1tsugr chol\_cat, tab bon

| Summary of total sugars (gm)

chol\_cat | Mean Std. dev. Freq.

------------+------------------------------------

1 | 117.80407 77.914108 2,514

2 | 116.05841 76.944908 1,532

3 | 106.70902 71.423402 244

------------+------------------------------------

Total | 116.54963 77.238602 4,290

Analysis of variance

Source SS df MS F Prob > F

------------------------------------------------------------------------

Between groups 27954.1146 2 13977.0573 2.34 0.0960

Within groups 25559369.4 4287 5962.06423

------------------------------------------------------------------------

Total 25587323.5 4289 5965.8017

Bartlett's equal-variances test: chi2(2) = 3.2523 Prob>chi2 = 0.197

Comparison of total sugars (gm) by chol\_cat

(Bonferroni)

Row Mean-|

Col Mean | 1 2

---------+----------------------

2 | -1.74566

| 1.000

|

3 | -11.095 -9.34939

| 0.097 0.237

PART 3

1. A picture containing diagram, plot, text, screenshot

   Description automatically generated
2. A picture containing screenshot, text, diagram, rectangle

   Description automatically generated

A picture containing text, screenshot, line, plot

Description automatically generated

reg lbxtc dr1ttfat

Source | SS df MS Number of obs = 4,290

-------------+---------------------------------- F(1, 4288) = 5.03

Model | 8466.1042 1 8466.1042 Prob > F = 0.0249

Residual | 7211258.76 4,288 1681.73012 R-squared = 0.0012

-------------+---------------------------------- Adj R-squared = 0.0009

Total | 7219724.86 4,289 1683.31193 Root MSE = 41.009

------------------------------------------------------------------------------

lbxtc | Coefficient Std. err. t P>|t| [95% conf. interval]

-------------+----------------------------------------------------------------

dr1ttfat | -.0316887 .0141235 -2.24 0.025 -.059378 -.0039994

\_cons | 195.5538 1.288524 151.77 0.000 193.0277 198.08

In the given regression model, the dependent variable is "lbxtc" and the independent variable is "dr1ttfat". The regression results are as follows:

The coefficient of "dr1ttfat" is -0.0317, which indicates that for every one-unit increase in "dr1ttfat", there is a decrease of 0.0317 units in "lbxtc", holding all other variables constant. The p-value for the coefficient is 0.025, which is less than 0.05, suggesting that the coefficient is statistically significant.

The intercept of the regression line is 195.5538, which means that when "dr1ttfat" is 0, the predicted value of "lbxtc" is 195.5538.

The R-squared value for the model is 0.0012, which means that only 0.12% of the variability in "lbxtc" can be explained by "dr1ttfat".

To assess the model fit, we can examine the residuals by plotting them against the predicted values. A good model fit would result in residuals that are randomly scattered around the horizontal line at zero. Additionally, we can check for linearity, normality, and constant variance of residuals.

In the results section, we can report that there is a statistically significant negative relationship between "dr1ttfat" and "lbxtc" (p=0.025). However, the R-squared value suggests that the model explains only a small amount of the variability in "lbxtc". The model fit can be improved by including other variables that may explain more of the variability in "lbxtc".

To plot the residuals against predicted values:

A picture containing screenshot, text

Description automatically generated

To test for linearity: If the relationship between the predictor variable and the outcome variable is linear, the scatter plot should show a straight line.

A picture containing screenshot, text

Description automatically generated

To test normality of residuals: If the residuals are normally distributed, the points on the QQ plot should fall along a straight line.

A picture containing text, screenshot, line, plot

Description automatically generated

To test for constant variance of residuals: If the variance of the residuals is constant, the scatter plot should show a random scatter of points around the horizontal line at zero.

A picture containing screenshot, text

Description automatically generated