```
* --- Query 1: The variable fincbtax is annual family income before tax. What are its mean,
* median, and standard deviation? Draw a histogram. Does it look normal (i.e., like a bell curve)?
* Does this distribution look skewed? Compute the standard error of the sample mean. What is the
margin of error for this estimate of the population mean? ---

* Summarize fincbtax
summarize fincbtax

* Calculate the median of fincbtax
centile fincbtax, centile(50)

* Draw a histogram of fincbtax
histogram fincbtax, normal title("Histogram of Family Income Before Tax (fincbtax)")

summarize fincbtax

* Calculate standard error of the sample mean (SE)
display "Standard Error of the Mean (SE): " r(sd) / sqrt(r(N))

* Calculate margin of error for 95% confidence interval
display "Margin of Error (95% CI): " 1.96 * (r(sd) / sqrt(r(N)))

* --- Query 2: Create a new variable called log_fincbtax that is equal to the natural log of
fincbtax.
* What is its mean, median, and standard deviation? Does its histogram look normal?
* Compute the standard error of the sample mean of this variable. What is the margin of error? ---

* Create the log_fincbtax variable
generate log_fincbtax = ln(fincbtax) if fincbtax > 0
label variable log_fincbtax "Log of annual family income before tax"

* Summarize log_fincbtax
summarize log_fincbtax
centile log_fincbtax, centile(50)

* Draw a histogram of log_fincbtax
histogram log_fincbtax, normal title("Histogram of Log(Family Income Before Tax)")

summarize log_fincbtax

* Calculate standard error of the sample mean for log_fincbtax
display "Standard Error of the Mean (SE) for log_fincbtax: " r(sd) / sqrt(r(N))

* Calculate margin of error for 95% confidence interval for log_fincbtax
display "Margin of Error (95% CI) for log_fincbtax: " 1.96 * (r(sd) / sqrt(r(N)))

* --- Query 3: Assuming log_fincbtax is normal, use the methods we learned in class to compute the
fraction
* of the sample that makes more than $100,000. That is, what is the probability that a normal random
variable
* with the mean and variance of log_fincbtax is greater than log(100000)? Compare this to the
fraction of the sample
* that has income higher than 100000. Did the normal approximation do a good job? ---

* Compute log(100000)
generate log_100000 = ln(100000)

* Compute the probability of log_fincbtax being greater than log(100000)
generate prob_gt_100000 = normal(r(mean) - log_100000)  // Normal CDF for log_fincbtax

* Calculate the actual fraction of the sample that has income greater than 100000
generate income_gt_100000 = fincbtax > 100000
summarize income_gt_100000
```

```
display "Fraction of sample with income > 100,000: " r(mean)

* Display normal approximation probability
display "Probability of log_fincbtax > log(100,000) (Normal Approximation): " 1 - normal(r(mean) - log_100000)

* --- Query 4: vehq is the number of vehicles owned by the household. What are its mean,
* median, and standard deviation? Draw a histogram. How would you describe the distribution?
* If it is a Poisson, what has to be true of the mean and variance? ---

* Summarize vehq and calculate the mean and variance
summarize vehq
display "Mean of vehq: " r(mean)
display "Variance of vehq (SD^2): " r(sd)^2

* Draw a histogram of vehq
histogram vehq, width(1) frequency title("Histogram of Number of Vehicles Owned (vehq)")

* Poisson condition: Mean ≈ Variance
display "For Poisson: Mean and Variance should be approximately equal."

* --- Query 5: Assume the distribution is a Poisson with lambda equal to the average of the observed
* mean and variance. Stata's poisson function takes two arguments (the lambda parameter and a value k)
* and computes the probability of a Poisson random variable being less than or equal to k.
* Use this function to compute the probability of a Poisson random variable with the lambda assumed
* above being 3 or fewer. How does this compare to the actual fraction of the sample that has 3 or fewer vehicles? ---

* Summarize vehq to get mean and variance
summarize vehq
scalar mean_vehq = r(mean)
scalar variance_vehq = r(sd)^2

* Calculate lambda as the average of mean and variance
scalar lambda = (mean_vehq + variance_vehq) / 2
display "Lambda (Poisson parameter): " lambda

* Compute the Poisson probability for k <= 3
scalar poisson_prob = poisson(lambda, 3)
display "Poisson Probability (k <= 3): " poisson_prob

* Calculate actual fraction of sample with vehq <= 3
generate vehq_leq3 = vehq <= 3
summarize vehq_leq3
scalar actual_fraction = r(mean)
display "Actual Fraction of Sample with vehq <= 3: " actual_fraction

* Compare Poisson probability and actual fraction
display "Comparison:"
display "Poisson Probability (k <= 3): " poisson_prob
display "Actual Fraction of Sample (vehq <= 3): " actual_fraction
```