1. This question uses daily data for the difference between the spot price and one-day ahead forward price in the California wholesale electricity market for the one hour period 5-6 p.m. for each day from April 1 1998 to December 31 1998.

diff = difference between spot and one-day ahead price in dollars per megawatt hour

(a) Does diff appear to be normally distributed? Explain your answer.

(b) What type of diagram would you use to visually display any outlying observations?

(c) What does the following Stata code do?

\[
generate \ y = (\text{diff} - \text{diff[_n-1]}) / \text{diff[_n-1]}
\]

(d) The original data were in a called electricity.dta Explain how you would read this data into Stata.

(e) If diff was actually normally distributed, what range of values would you expect 95% of the observations on diff to fall in, given the output on the last sheet?
2. For this question continue to use the information given at the end of the exam. Clearly state any details that you use along the way including, for tests, the hypothesis tested and the alternative hypothesis and your conclusion.

(a) Give a **95 percent** confidence interval for population mean price difference.

(b) Give a **90 percent** confidence interval for population mean price difference.

(c) The claim is made that the population mean price difference is zero. Test this claim at significance **level 0.05**.

(d) Give the mathematical formula for the p-value of the test in part (c).

(e) Suppose that a positive price difference one day means that the price difference the next is also likely to be positive. How would this effect your previous analysis? Explain.

(f) Suppose that instead of daily data from April 1998 to December 1998 we only had monthly data. How would this effect your previous analysis? Explain.
3. (a) Consider a simple random sample of size 4 with values 2, 8, 5, 1. Compute the sample mean, variance and standard deviation. **Show all workings.**

(b) Let \( X \) be the number of company mergers this year in the telecommunications industry. Suppose \( X = 0 \) with probability 0.4, \( X = 1 \) with probability 0.2 and \( X = 2 \) with probability 0.4. Compute the mean, variance and standard deviation of \( X \). **Show all workings.**

(c) Suppose for \( X \sim (100, 20^2) \) we form 1000 samples of size 50 and obtain 1000 sample means \( \bar{x} \). What approximately do you expect the average of the \( \bar{x} \) to equal? What approximately do you expect the standard deviation of the \( \bar{x} \) to equal?

(d) Consider a random variable that takes values \( X = 1 \) with probability 0.8 and value \( X = 0 \) with probability 0.2. State how you would generate a random sample of size 50. Give a step-by-step method. (You do not need to give exact Stata commands.)
Multiple Choice Questions (1 point each)

1. Data on number of doctor visits in 2012 for a sample of 192 individuals is an example of
   a. categorical cross-section data
   b. numerical cross-section data
   c. categorical time series data
   d. numerical time series data
   e. none of the above

2. The symmetry statistic is approximately
   a. $s^3$
   b. $\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^3$
   c. $\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^3 / s^3$
   d. none of the above

3. For data from a simple random sample, $(\bar{x} - \mu)/(s/\sqrt{n})$ is a realization of a random variable that is distributed as
   a. $t_{n-1}$ approximately if $n$ is large
   b. $t_{n-1}$ if $X$ is normal
   c. $t_{n-1}$ always
   d. all of a., b., and c.
   e. both a. and b.

4. For a simple random sample, $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ is
   a. standardized to have mean 0 and variance 1 always
   b. normally distributed as $n \to \infty$
   c. neither of the above
   d. both of the above

5. A key feature of the approach to statistical inference from simple random samples is that
   a. the population mean is unchanging but the sample mean changes
   b. the sample mean is unchanging but the population mean changes
   c. both the population mean and the sample mean change
   d. both the population mean and sample mean are unchanging
SOME USEFUL FORMULAS

Univariate Data

\[ \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{and} \quad s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \]

\[ \bar{x} \pm t_{\alpha/2,n-1} \times (s_x/\sqrt{n}) \quad \text{and} \quad t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \]

ttail(df, t) = Pr[T > t] where T \sim t(df)

t_{\alpha/2} such that Pr[|T| > t_{\alpha/2}] = \alpha is calculated using invttail(df, \alpha/2).

Bivariate Data

\[ r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \times \sum_{i=1}^{n} (y_i - \bar{y})^2}} = \frac{s_{xy}}{\sqrt{s_{xx} \times s_{yy}}} \]
[Here \( s_{xx} = s_x^2 \) and \( s_{yy} = s_y^2 \),]

\[ \hat{y} = b_1 + b_2 x_i \]
\[ b_2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \]
\[ b_1 = \bar{y} - b_2 \bar{x} \]

\[ \text{TSS} = \sum_{i=1}^{n} (y_i - \bar{y})^2 \]
\[ \text{ErrorSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]
\[ \text{RegSS} = \text{TSS} - \text{ErrorSS} \]

\[ R^2 = 1 - \frac{\text{ErrorSS}}{\text{TSS}} \]

\[ b_2 \pm t_{\alpha/2,n-2} \times s_{b_2} \]

\[ t = \frac{b_2 - \beta_{20}}{s_{b_2}} \]

\[ s_{b_2}^2 = \frac{s_e^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad s_e^2 = \frac{1}{n-2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

\[ y|x = x^* \in b_1 + b_2 x^* \pm t_{\alpha/2,n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} + 1} \]

\[ \text{E}[y|x = x^*] \in b_1 + b_2 x^* \pm t_{\alpha/2,n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \]

Multivariate Data

\[ \hat{y} = b_1 + b_2 x_{2i} + \cdots + b_k x_{ki} \]

\[ R^2 = 1 - \frac{\text{ErrorSS}}{\text{TSS}} \quad \bar{R}^2 = R^2 - \frac{k-1}{n-k}(1 - R^2) \]

\[ b_j \pm t_{\alpha/2,n-k} \times s_{bj} \quad \text{and} \quad t = \frac{b_j - \beta_{j0}}{s_{bj}} \]

\[ F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \quad \text{and} \quad F = \frac{(\text{SSE}_e - \text{SSE}_u)/(k-g)}{\text{SSE}_u/(n-k)} \]

Ftail(df1, df2, f) = Pr[F > f] where F is \( F(df1, df2) \) distributed.

\[ F_{\alpha} \] such that Pr\[F > f_{\alpha}\] = \alpha is calculated using invFtail(df1, df2, \alpha).
. sum diff, detail

<table>
<thead>
<tr>
<th>Percentiles</th>
<th>Smallest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1%</td>
<td>-100.7092</td>
</tr>
<tr>
<td>5%</td>
<td>-36.7719</td>
</tr>
<tr>
<td>10%</td>
<td>-20.9986</td>
</tr>
<tr>
<td>25%</td>
<td>-8.998913</td>
</tr>
<tr>
<td>50%</td>
<td>-2.2891</td>
</tr>
<tr>
<td>75%</td>
<td>2.870001</td>
</tr>
<tr>
<td>90%</td>
<td>19.2143</td>
</tr>
<tr>
<td>95%</td>
<td>70.69516</td>
</tr>
<tr>
<td>99%</td>
<td>175.58</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Largest</th>
</tr>
</thead>
<tbody>
<tr>
<td>75%</td>
</tr>
<tr>
<td>90%</td>
</tr>
<tr>
<td>95%</td>
</tr>
<tr>
<td>99%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean</th>
<th>Mean Std. Err.</th>
<th>[95% Conf. Interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>diff</td>
<td>1.421385</td>
<td>2.145666</td>
</tr>
</tbody>
</table>

. display "t_273,.025 = " invttail(273,.025) " t_273,.05 = " invttail(273,.05) t_273,.025 = 2.5939578 t_273,.05 = 1.6504543

. display "t_273,.005 = " invttail(273,.005) " t_273,.01 = " invttail(273,.01) t_273,.005 = 2.5939578 t_273,.01 = 2.3400846

. display "t_273,.025 = " invttail(273,.025) " t_273,.05 = " invttail(273,.05) t_273,.025 = 1.9686916 t_273,.05 = 1.6504543