Stat 231 Homework Assignments/Spring 2010

HW1/ Due 1/14/10

1. Do Devore Chapter 1 Problem 42.

2. Do Devore Chapter 1 Problem 51 "by hand." That is, do the arithmetic required to compute the sample variance and sample standard deviation. Check your computations using a good scientific calculator (you need to know how to compute means and standard deviations using your calculator). Be sure to write correct units for $s^2$ and $s$.

3. Do Devore Chapter 1 Problem 64 "by hand." That is, do the arithmetic required to compute the two sample means and two sample standard deviations. Check your computations using a good scientific calculator (you need to know how to compute means and standard deviations using your calculator). Be sure to write correct units for the $\bar{x}$ and $s$ values.

4. Load the data from Devore's Example 1.9 into JMP. Use the Analyze>Distribution function to get and print out a JMP report that includes
   a) a histogram matching the one in the text as closely as is possible (run the data axis from 1 to 19 and use increments of 2),
   b) a quantile box plot,
   c) a stem and leaf plot, and
   d) the sample mean and standard deviation for these data.

5. Load the data from Devore Chapter 1 Problem 75 into JMP. Use the Tables>Stack function to get the data values into a single column and the labels into another one. Then make and print out the comparative (quantile) box plots Devore asks for in two different ways.
   a) Use the Analyze>Distribution function plotting Data by Label and then using horizontal layouts to stack the JMP reports over one another.
   b) Use the Graph>Graph Builder function, plotting Data on the "y" axis and Label on the "x" axis, and then adding the box plots.

HW2/Due 1/21/10

Do Devore Chapter 2 Problems 1,2,3,8,11,13,14,15,19,20,26,45,48,52,53,54,67.

HW3/Due 1/28/10

HW4/Due 2/4/10

1. Use JMP and make "spike graphs" to the same scale for probabilities from all of the following discrete distributions:
   a) the Binomial distributions for \( n = 5 \) and \( p = 0.1, 0.2, 0.5, 0.8 \),
   b) the Geometric distributions for \( p = 0.1, 0.2, 0.5 \), and
   c) the Poisson distributions for \( \lambda = 0.5, 1, 2, 4 \).

Make these plots of probability mass functions \( f(x) \) vs \( x \) for \( x = 0, 1, 2, \ldots, 19 \). You can do this by
   a) adding 20 rows to an initial data table and creating a first column (label it "x")
      using a formula that is Row() - 1 (this gives "row number - 1") (you'll find "Row" under the grouped "Row" functions menu in the "Formula" menu),
   b) adding columns to the data table for each of the distributions indicated above (you'll find appropriate probability mass functions under the grouped "Discrete Probability" functions menu in the "Formula" menu),
   c) using the Graph Builder function with \( x \) on the horizontal axis and the various probability function values on the vertical axes. If you right-click on each graph, you'll be able to remove the "smoother" and add "bars" to make the desired plots.

Turn in a plot of these graphs.

2. Use JMP and simulate 1000 values from each of the distributions
   a) Binomial for \( n = 5 \) and \( p = 0.2 \),
   b) Geometric for \( p = 0.2 \), and
   c) Poisson for \( \lambda = 2 \).

Then make (and turn in) histograms for the each of the 3 sets of values and say how they compare to the corresponding probability mass functions you plotted above.

You may do the simulation by adding 1000 rows to an initial data table and creating 3 different columns using functions from the "Random" group of functions in the "Formula" menu. Once the values have been simulated into the columns, you may summarize them exactly as you summarized empirical data in HW1.

---

1 JMP does not have a simple Geometric probability function. You may either build your own formula for it or use the more general "Negative Binomial" probability function. Devore discusses the negative binomial distributions in Section 3.5. They are distributions for the trial at which the \( r \)th success occurs in a sequence of Bernoulli trials. For better or worse, the JMP is slightly different, being the distribution for this number of trials minus 1 (i.e. the distribution for the number of trials before the \( r \)th success. Thus, if you want to use the JMP function, you will need to evaluate at \( p, 1, \) and \( x - 1 \), e.g. use Neg Binomial Probability \( (.1, 1, x - 1) \) for the first case.

2 JMP allows you to simulate from a simple Geometric distribution, BUT it calls the number of trials before the first success. So to get something fitting our class notation,
3. Do Devore Chapter 3 Problems 48,49,50,51,60.

4. In the context of Problem 3.48, let
   \( Y = \) the number of boards inspected to find a defective one
   
   a) What is the name of the distribution of \( Y \)? (State too the value of its parameter.)
   b) Find \( P[Y = 4] \) and \( P[1 \leq Y \leq 3] \).
   c) Find \( EY \) and \( VarY \).

5. In the context of Problem 3.50, let
   \( W = \) the number of calls that must be received in order to get a fax
   
   a) What is the name of the distribution of \( W \)? (State too the value of its parameter.)
   b) Find \( P[W = 4] \) and \( P[W > 4] \).
   c) Find the mean and standard deviation of \( W \).

6. Do Devore Chapter 4 Problems 1,3,4,11,12,13.

**HW5/Due 2/11/10**

1. Do Devore Chapter 4 Problems 28,29,30,32,34,42,43,46,47,53,54,59,61,69,73.

2. Do Devore Chapter 5 Problems 1,3,4,9 (a through d),10 (b and c, the joint density is 1 on the square where both \( X \) and \( Y \) are between 5 and 6 and 0 otherwise),12 (in b, only compute the two marginal densities),18.

**HW6/Due/2/18/10**

1. Do Devore Chapter 5 Problems 9e, 12c (answer the question about independence),13,15,25,26,27,31,32,77.

Simulation is a simple way to get approximate answers to some probability problems.
We've already done a bit of simulation in HW4. Here we'll use it to get quick/approximate answers to some of the problems you've worked out "by hand" using calculus. This is easy to do in JMP, if what is needed is some probability defined in
terms of independent random variables that themselves have "standard" distributions (that JMP knows how to simulate from).

you will need to add 1 to values generated directly by JMP. That is, the formula to use for the column where you put the Geometric(.2) random values is

Random Geometric(.2)+1.
2. Return to the scenario of Problem 10 of Chapter 5 of Devore. Open a new JMP data table and add 1000 rows. Make a column labeled "X" that has in it 1000 simulated values of Annie's arrival time. You may do that using the "Random Uniform" function to make simulated values uniform on the interval (0,1) and then adding 5 to them. (That is, the formula to use for the column is Random Uniform( ) + 5.) Do the same to make a column labeled "Y" for Alvie's arrival times. Then make a third column (call it "Z" or anything else you like) with a formula |X−Y|. (You can find the absolute value function in the "Numeric" function group.) This new column now holds simulated values of the difference in arrival times for Alvie and Annie. You may now use the Analyze Distribution function to see empirical properties of these simulated values that serve as approximations to theoretical properties of the distribution of |X−Y|.

a) What, approximately, is the mean absolute difference arrival times (E|X−Y|)? (Find the sample mean of your simulated values of Z.)
b) Add the "CDF Plot" to the JMP report display, and use the crosshair tool (button on the JMP task bar) to read off the plot the answer to part c) of problem 10.

Attach to your homework a printout of the JMP report with the approximate value of the mean absolute difference in arrival times circled and crosshairs sketched by hand in on the CDF Plot showing your answer to b).

3. Return to the scenario of Problem 13 of Chapter 5 of Devore. Use simulation to get an idea of the shape of the distribution of X+Y (simulate values and make a histogram of the simulated values). Does this distribution appear to be exponential like those of X and Y? Use the method indicated above (involving CDF Plot) to get approximate answers to parts c) and d) of problem 13. Turn in the JMP printout with your homework.

4. Consider the scenario of Problem 15 of Chapter 5 of Devore. Suppose for sake of argument that λ=1. The lifetime of the system can be written as

Y = minimum( X₁, maximum( X₂, X₃ ))

Use simulation to get an idea of the shape of the distribution of Y and the mean lifetime EY. Turn in the JMP printout with your homework. (The minimum and maximum functions are in the "Statistical" group of functions and you will need to put commas between the things you wish to take a minimum or maximum of.)

HW7 Due 3/4/10

1. Use simulation to get an idea of the nature of the distribution of \( \bar{X} \) for independent identically distributed random variables that are individually Uniform(0,1). Simulate 1000 values of \( X₁, X₂, \ldots, X₁₀ \) all Uniform(0,1). Then compute 1000 values for \( \bar{X} \) based on \( n = 2, 5, \) and \( n = 10 \). (The "Mean" function in the group of "Statistical" functions can be used, or you can simply use the "hand" formula for the sample average.) Get JMP reports providing histograms for \( X₁ \) and your sample
averages for the 3 different sample sizes. Change the axes on these so the same horizontal scale is used on all 4 histograms (so that they may be easily compared). Display the 4 reports stacked vertically so the histograms are lined up in order, \( n = 1 \) through \( n = 10 \). (You may do this by preparing all the columns to be summarized at one time, making a single use of the Analyze \( \rightarrow \) Distribution dialogue box to do all the summarizations at once, and then employing the Stack option under the top red triangle of the Distributions report. Then you may change the horizontal axis settings on the histograms to make them match.)

For each of the 4 cases, also use the CDF Plot idea from the previous homework to approximate the probability that \( \bar{X} \) is within \( 1/\sqrt{12n} \) (the theoretical standard deviation of \( \bar{X} \)) of 0.5 (the theoretical mean of \( \bar{X} \)). How do these compare to 0.6826 (the value of \( P[-1 < Z < 1] \))?

2. Redo Problem 1, using \( \text{Exp}(1) \) random variables. For the last part of the problem, note that the theoretical standard deviation is \( 1/\sqrt{n} \) and the theoretical mean is 1. Use these values in place of the values \( 1/\sqrt{12n} \) and \( 1/2 \) employed in problem 1.

3. Do Devore Chapter 5 Problems 46,47,48,52,53,54,93.

4. Do Devore Chapter 7 Problems 1,3,4,12,14,18,19,23,28,29,30,32,34,35,38,42,43,44,45.

5. Do a small simulation to illustrate the intent of confidence and prediction intervals for a normal process. That is, do the following. First simulate 20 sets of values (20 rows) of independent standard normal variables \( X_1, X_2, \ldots, X_5 \) and \( X_{\text{new}} \). Then make new columns containing \( \bar{X} \) and \( S \) based on only the \( n = 5 \) observations \( X_1, X_2, \ldots, X_5 \). From these values of the sample mean and standard deviation make new columns that contain
   - upper and lower two-sided 90% two-sided confidence limits for \( \mu = 0 \) based on display (7.15) of Devore and the \( t_4 \) distribution (compute 2 new columns),
   - upper and lower two-sided 90% prediction limits for \( X_{\text{new}} \) based on display (7.16) of Devore and the \( t_4 \) distribution (compute 2 new columns), and
   - upper and lower two-sided 90% confidence limits for \( \sigma = 1 \) based on page 279 of Devore.

Turn in a copy of your data table with this HW set. For each of your samples of size \( n = 5 \) (for each of your 20 rows), indicate with a handwritten check mark on the row those samples for which the confidence limits for \( \mu \) are successful in locating the parameter \( \mu = 0 \), indicate with a star on the row those samples for which the confidence limits for \( \sigma \) are successful in locating the parameter \( \sigma = 1 \), and indicate with an "X" on the row those samples for which the prediction limits for \( X_{\text{new}} \) are successful in locating the 6th standard normal observation \( X_{\text{new}} \).
HW8/ Due 3/11/10

1. Return to the small simulation that was part 5. on the last assignment. Add 2 more columns to the table there that contain 95% two-sided tolerance limits for 99% of the standard normal distribution (99% of all additional observations from the process). (These are $\bar{X} \pm 6.634 \cdot S$ for this sample size of $n = 5$.) Then from these two columns, compute in another column the actual fraction of the standard normal distribution caught between these two limits. (Note that this is $\Phi(\bar{X} + 6.634 \cdot S) - \Phi(\bar{X} - 6.634 \cdot S)$ and that the standard normal cumulative distribution function $\Phi$ is available as the Normal Distribution function under the Probability group of functions.) Finally, Indicate with a double "XX" on the row, those samples for which the tolerance intervals are successful in locating at least 99% of the standard normal population. If you made not just 20, but rather many many rows of simulated values and tolerance limits, about what fraction of those rows would you expect to have "XX" on them?

2. Do Devore Chapter 4 Problems 87, 91. (For Problem 91, use the Fit Distribution option in JMP to make the plot. Under the top red triangle, choose Continuous Fit $\to$ Normal and then under the red triangle next to Fitted Normal activate Diagnostic Plot. This will produce the desired normal probability plot. Deactivate the "Confidence Limits" under the red triangle next to Diagnostic Plot, leaving only a fitted line on the plot in addition to the plotted points. Turn in a copy of your JMP report with the homework.)

3. Do Devore Chapter 8 Problems 1,2,3,6,11,14,15,16,19 (a and b only). These problems deal more or less with the general logic/"theory" of hypothesis testing.

4. Do Devore Chapter 8 Problems 20,24,27,31,35,36 (a and b only),49. These problems concern specific (more or less realistic) applications of hypothesis testing.

(Vardeman will send you an appropriate JMP file so that you don't have to key in the values for Problem 27.)

HW9/Due 3/25/10

1. Devore Chapter 8 Problem 80.

2. Devore Chapter 9 Problems 2,4,8,19,20,22 (and make a 95% two-sided confidence interval for the difference in mean shear strengths), 30, 36, 37, 47 (and make 95% two-sided confidence limits for the difference in the proportions of switchers), 52 (both test a hypothesis via computation of a p-value and make 95% two-sided confidence limits for the difference in proportions of marketable bulbs), 57, 58, 63 (make a two-sided interval and state limits for the ratio of standard deviations as well ones for the ratio of variances), 64.
HW10/Due 4/8/10

1. Below is a small set of fake $(x, y)$ data. Show the "hand" calculations necessary to:
   a) Find the least squares line through the data, the sample correlation between $x$ and $y$, the sample correlation between $y$ and $\hat{y}$, and $R^2$.
   b) Assuming the simple linear regression model
      \[ y_i = \beta_0 + \beta_1 x_i + \epsilon_i \]
      where the $\epsilon_i$ are independent $\text{Normal}(0, \sigma^2)$ random variables.
      i) Find $SSE$ and give a single number estimate of $\sigma^2$.
      ii) Use a $t$ test with $\alpha = 0.05$ to test $H_0: \beta_1 = 0$ vs. $H_a: \beta_1 \neq 0$.
      iii) Give a 90% two-sided confidence interval for the increase in average $y$ that accompanies a unit increase in $x$.
      iv) Give a 90% two-sided confidence interval for the average value of $y$ when $x = 2$.
      v) Give a 90% prediction interval for the next value of $y$ when $x = 2$.
      vi) Make an ANOVA table for the data and find the value of an $F$ statistic that can be used to test the hypothesis in ii).
   c) Get a JMP report that you can use to check your hand calculations. Annotate the printout marking on it the locations of all the pieces needed to do the things above and turn it in with your hand calculations.

\[
\begin{array}{c|cccc}
  x & -1 & 0 & 1 & 2 & 3 \\
  \hline
  y & 4 & 4 & 3 & 3 & 2 \\
\end{array}
\]

2. Do Devore Chapter 12 Problems 1b) and c), 7, 8, 14, 19, 23, 30 (this is a "theoretical" problem about the simple linear regression model, not a data analysis problem), 38 (do both $t$ and $F$ tests), 52 (also make and interpret 95% two-sided confidence limits for $\sigma$ in this model). You may use JMP to facilitate any of these calculations, but if you do, attach a carefully annotated printout of the relevant JMP report.

HW11/Due 4/16/10

1. Do Devore Chapter 13 Problems 3 (also normal-plot the standardized residuals and comment on the implications of the appearance of the plot), 14, 19, 31a and c (Also, use JMP "Fit Y by X" to obtain a plot of the fitted quadratic, confidence limits for mean response, and prediction limits for the next response. Then use JMP "Fit Model" to fit the quadratic and save standardized residuals for the quadratic fit. Plot these standardized residuals for the quadratic fit versus volume hourly space velocity and normal-plot them, commenting on any interesting features of the plots.), 35, 37, 38, 42 (get your own JMP report from "Fit Model" to use as a basis for answering these questions), 47 (get your own JMP report from "Fit Model" to use as a basis for answering these questions).
HW12/Due 4/23/10 (This is the day of Exam 3)

Below are the raw data taken from Example 13.18 of Devore. The response $y$ is wear life (on journal bearings) in hours for a solid film lubricant, $x_1$ is the load applied in 1000 psi and $x_2$ is a speed in rpm.

<table>
<thead>
<tr>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>20</td>
<td>300.2, 310.8, 333.0</td>
</tr>
<tr>
<td>6</td>
<td>20</td>
<td>99.6, 136.2, 142.4</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>20.2, 28.2, 102.7</td>
</tr>
<tr>
<td>3</td>
<td>60</td>
<td>67.3, 77.9, 93.9</td>
</tr>
<tr>
<td>6</td>
<td>60</td>
<td>43.0, 44.5, 65.9</td>
</tr>
<tr>
<td>10</td>
<td>60</td>
<td>10.7, 34.1, 39.1</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>26.5, 22.3, 34.8</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>32.8, 25.6, 32.7</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>2.3, 4.4, 5.8</td>
</tr>
</tbody>
</table>

Use JMP to help you do all that follows and attach your annotated JMP report(s) to your work.

a) Fit the model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

to the data. Plot standardized residuals against $x_1$, against $x_2$, and against $\hat{y}$. What do these indicate about the appropriateness of this model?

b) One way to try to improve the fit to these data is to fit a model with additional terms. Fit the general 2-variable quadratic model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \epsilon$$

to these data. Plot standardized residuals against $x_1$, against $x_2$, and against $\hat{y}$. What do these indicate about the appropriateness of this model?

c) Compare the fits of the two models in a) and b) on the basis of the appearance of the plots, $R^2$ values and $C_p$ values. Is either a really adequate description of the data? Explain.

d) Temporarily adopt the model from (b) and test $H_0: \beta_3 = \beta_4 = \beta_5 = 0$ in this model using $\alpha = .05$.

Devore points out that the standard method of analysis of data such as these is to take logs of both the predictor variables and the response variable. Accordingly, create new variables in JMP $y' = \ln(y)$, $x_1' = \ln(x_1)$, and $x_2' = \ln(x_2)$.

e) Fit the model
to the data. Plot standardized residuals against $x_1'$, against $x_2'$, and against $\hat{y}'$. What do these indicate about the appropriateness of this model?

Regardless of the outcome of your residual analysis from e), henceforth use the model of e).

f) Plot values of $y'$ against $x_1'$. On the same set of axes, plot the 9 sample means of $y'$ (for the different $(x_1', x_2')$ combinations) against $x_1'$, connecting with a "curve" composed of two line segments those means corresponding to a given value of $x_2'$. Finally, on this plot the 9 values of $\hat{y}'$ against $x_1'$, and make similar "curves" for the three different $x_2'$ values in the data set. Based on this plot, comment on how well the model from e) appears to describe reality.

g) Test $H_0: \beta_1 = \beta_2 = 0$ in the model of e) (using the overall F test and $\alpha = .05$).

h) Do a formal test for lack of fit to the model from e) using $\alpha = .05$.

i) Give 95% two-sided confidence intervals for $\beta_1$ and $\beta_2$ in the model from e).

j) Give a 95% two-sided confidence interval for $\mu_{y|x_1=6,x_2=60}$. Then give a two-sided 95% prediction interval for the next $y'$ when $x_1 = 6$ and $x_2 = 60$. Exponentiate the end points of your prediction interval to produce a prediction interval for the next $y$ when $x_1 = 6$ and $x_2 = 60$.

k) Repeat j) but using the set of conditions $x_1 = 8$ and $x_2 = 80$. (Note that you will have to specifically tell JMP to give you the appropriate predicted value and standard deviation of fit for these conditions.)

**HW13/Not to be Collected**

In an ISU engineering research project, so called "tilt table tests" were done in order to determine the angles at which certain vehicles experience lift-off of the "high side" set of wheels and begin to roll over on their sides. So called "tilt table ratios" (which are the tangents of angles at which lift-off occurred) were measured for 4 different vans with the following results.

<table>
<thead>
<tr>
<th>Van #1</th>
<th>Van #2</th>
<th>Van #3</th>
<th>Van #4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.096, 1.093, 1.090, 1.093</td>
<td>.962, .970, .967, .966</td>
<td>1.010, 1.024, 1.021, 1.020, 1.022</td>
<td>1.002, 1.001, 1.002, 1.004</td>
</tr>
</tbody>
</table>
(Notice that Van #3 was tested 5 times while the others were tested 4 times each.) Vans #1 and #2 were minivans and Vans #3 and #4 were full size vans.

a) If one is to apply the methods of Chapter 10 of Devore in the analysis of these \( r = 4 \) samples, what model assumptions have to be made? With the small sample sizes available here, only fairly crude checks on the appropriateness of the assumptions are possible. One possibility is to compute and normal-plot residuals and/or standardized residuals. Do this and comment on the appearance of your plots. (Normal plot both residuals and standardized residuals.)

b) Compute a pooled estimate of the standard deviation based on these 4 samples. What is \( s_p \) supposed to be measuring in this example?

c) Use your value of \( s_p \) and make 95% two-sided confidence intervals for the 4 different mean tilt table ratios for the 4 vans, \( \mu_1, \mu_2, \mu_3, \) and \( \mu_4. \)

d) Confidence intervals for the differences between particular pairs of mean tilt table ratios are of the form \( \bar{y}_i - \bar{y}_j \pm \Delta \) for appropriate values of \( \Delta \). Use \( s_p \) and find values of \( \Delta \) if 95% two-sided intervals are desired, first for pairs of means with samples of size \( n_i = n_j = 4 \) and then for pairs of means where one sample size is \( n_i = 4 \) and the other is \( n_j = 5. \)

e) It might be of interest to compare the average of the tilt table ratios for the minivans to that of the full size vans. Accordingly, give a 95% two-sided confidence interval for the quantity

\[
\frac{1}{2} (\mu_1 + \mu_2) - \frac{1}{2} (\mu_3 + \mu_4)
\]

f) Show the "hand" calculations necessary to make an ANOVA table for testing

\( H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4. \) Using the F table in Devore, what can be said about the \( p \)-value for this test?

g) Use \textit{JMP} to get an ANOVA table for checking your work in part f). What \( p \)-value does \textit{JMP} give you? What does this say about the mean tilt table ratios?