Office Hours: 4:30 – 5:30 Wednesday, Thursday, and by appointment. To be held in Room 106. Office hours provide you with an opportunity for personal discussion with me concerning course related problems such as homework, clarification of classroom discussion, test grading etc. If the formal hours are unsuitable for you, please make an appointment with me at a mutually agreeable time. I strongly urge you to take advantage of these hours. My graduate teaching assistant, ShaoBo Li (lis6@mail.uc.edu) will also have office hours in his office in Room 534 Lindner Hall that I will announce early in the semester.

Historical Information: This is the 5th time this course has been offered in its current format, although it has been offered in a different format since 1992. It owes its existence to the move from quarters to semesters a while ago. The topics are combined in part from two previously existing courses that seem to go well together, although other courses could have been combined with equally logical validity. Below, the goals and descriptions make clear that the topics come from, 1) statistical methods and 2) basic linear regression models. I will regard these two related subjects as giving rise to essentially two separate modules, each with its own homework, tests, and grades (which will be combined at the end). This is reflected in the rest of this document. I intend to devote the first 9 or 10 weeks to the statistical methods module (called module 1 in the sequel), and the remaining 4 or 5 weeks to the basic linear regression models module (called module 2 in the sequel). Regression modeling will then form the core of the sequel, BANA 7042, offered in the first flex term of the new year. BANA 7041 is the prerequisite.

Computational Aspects of this Course: The main computational tool we will use, namely SAS, will be available free to all students by using UCvlabs. You can learn how to gain access to the computers by following http://business.uc.edu/technology/services/ucvlabs.html. You will find that you can use SAS on any computer at anytime and anywhere by using UCvlabs. As a second alternative, a license for SAS that resides on your own computer can be purchased from the bookstore for some amount ($71 last year). But, it has the disadvantage that it expires and a renewal costs more money. I don't know much more about this option, so check with the bookstore. Finally, SAS is available in the LCB computer lab on every machine. But this requires you to be present to use them. Yet another option is described in the section of this syllabus entitled Texts.

This is strictly a SAS-centric course. I am always asked if I will accept homework and tests where computations are done using some other software such as SPSS, R, etc. My answer is “NO”. If you have a need to use and/or learn other statistical software that is fine with me if you want to gain experience on how to use these other methods or to check your SAS answers. But it is not acceptable for submitted work. SAS knowledge is an integral part of this course. Occasionally, you may want to use Excel to organize data or to check on the SAS answers. That is fine. In fact, you can use any software for this purpose. Mastery of SAS is part of this course. If you cannot do a problem using SAS, it is acceptable to use the mathematical formulas to solve problems “by hand”, but that is a pity. I cannot think of a problem yet that could not use some feature of SAS to help you solve. I know some of you will find learning how to use SAS to do statistics difficult. My hope is that your task in learning SAS will be made easier by imitating my practice problems,
or the demonstrations in my class, or the collection of SAS demonstrations in the SAS folder of Blackboard, or the examples you might find by making a sincere effort to search the Internet. Keep trying. I know you will succeed.

Scheduling Two Different Sections: There are two sections of this course, 001 (Wednesday) and 002 (Thursday). On all of your submitted work, please indicate which section you/your group reside by including the appropriate section number clearly written on the front of your submitted paper next to your names. This year, recording of the lectures in the Wednesday and Thursday classes will be made available to both sections shortly after (usually a day) the class is complete.

Section 001 and 002 are officially scheduled to meet on different days and for a different number of times due to Thanksgiving (Thursday 11/24), as well as the reading-day (Thursday 10/13). This is not reasonable for two sections of the exact same course, in my opinion. That is, 15 meeting for section 001 and only 13 meeting for section 002. To remedy this disparity, I am taking special measures to equalize the lecture content and to equalize. Specifically, the Wednesday Section 001 class will be held as scheduled on 10/12. But, the Thursday Section 002 that is cancelled for reading-day on 10/13 will be assigned as MANDATORY video viewing of the previous day’s lecture for that week. The Wednesday 11/23 class will not be held since the Thursday 11/24 is cancelled. In this way, both sections will have a total of 14 lectures with only one mandatory video lecture. This video-viewing assignment is very important and is mandatory for all students concerned. By the way, as long the room capacity is not violated, students from either section can attend either class.

Recording of Classes: There are two sections of this course, 001 and 002. Recording of the lectures in both of the classes will be made available to both sections shortly after (usually a day) the class is complete. We are using Echo 360 to record the lectures in each of the two classrooms. A tab should be available in the Meta_Levy_238 Blackboard site that will be the home for both sections. You should find the videos of all lectures available when you click that tab there. You can view any ones you like, but I assume you will want to view the ones corresponding to the dates of your lectures. The exception will be for the special measures I have described due to the cancelled class for which mandatory video viewing have been assigned as describe in the above section.

Tests: For module 1, there will be one midterm exam. This will be a take-home disseminated around the sixth week of class and due one week later in class. There will be a two part comprehensive final for module 1; one part will be a take-home exam disseminated on the last class of the first module and due one week later (during the first week of module 2). The second part of the Module 1 final will be short-answer in-class and will be combined with the entire course comprehensive in-class portion of the Module 2 final. This completely course-comprehensive mandatory Final Exam for both sections will be given as a Block Final Exam and will be scheduled by the Registrar’s office in a room and a time to be announced later, so plan up front for that when these details are revealed. For module 2, there will be a two part final exam as well. The take-home part will be disseminated on the last week of class and due on or before Thursday, Dec. 8 at 6:30 PM in room 414 Zimmer. The short answer in-class part for module 2 will be the second half of the course comprehensive final combined with the module 1 short answer test as described above as a Block Final Exam. If you are forced to miss submitting the midterm for a documented reason beyond your own control, the comprehensive course short answer final exam score will be substituted for the missing midterm exam score. If, for a similar reason, you are forced to delay submitting the completed copy of the take-home final, a special arrangement with me must be made. In all cases, I will be the judge of the severity and validity of the excuse. If this protocol is not followed, the exam score will be counted as zero. You must contact either the OBAIS department administrative assistant or me before missing or delaying submission of the midterm (or final). Otherwise, the exam score will be counted as zero. Take-home exams will be subject to the group-work policy stated later in this document. In-class exams are individual efforts. The same policy regarding missed exams above holds for module 2. Cheating will be subject to the university policy on student conduct.

<table>
<thead>
<tr>
<th>FINAL EXAM FOR…</th>
<th>TYPE</th>
<th>WHEN</th>
<th>WHERE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mod 1</td>
<td>Take home - group</td>
<td>Week 9 or 10, due week later</td>
<td>My office or classroom</td>
</tr>
</tbody>
</table>
Mod 1* In class – individual TBA TBA
Mod 2 Take home - group 14th week, due last day of class or before My office or classroom
Mod 2* In class - individual TBA TBA

* Complete course comprehensive, i.e., Mod 1 and 2 take in-class exams combined

**Homework:** There are two types of assignments, type 1 and type 2.

1. Practice problem assignments that are listed in the Practice Problem folder of Blackboard. Those practice problems for the early part of the course in Module 1 are simply a selection of the numbered Examples from the text by Snedecor and Cochran. These are not to be handed in. They are for practice only and to be attempted individually or with your group. I will have my solutions posted there too (mostly using SAS) in another folder. You are expected to review a textbook practice problem, corresponding to a day’s lecture, within one week of that lecture. These textbook solutions are there as a resource to help you learn statistical methods and how to use SAS to perform those statistical analyses. It turns out that the solutions using SAS to solve the practice problems will be amazingly helpful for attacking the hand-in problems and more importantly the exam problems.

2. Hand-in problems: These form a collection of problems with SAS solutions. These are being made available to you in a separate folder in Blackboard. For the most part, these will have accompanying answers using SAS. This is a second resource for you to learn how to do statistical analyses using SAS. These problems are designed to increase your knowledge of statistical methods and the proper use of SAS. Optimally you should try to solve these **without recourse to my solutions** to get the maximum benefit. But I am fully aware of the time pressures facing you. So, at a minimum you (or your group) must reproduce the SAS codes in these solutions to gain proficiency performing statistical analyses using SAS. But as per #4 of the Assurance of Learning Objective, I expect more than just a simple answer, copying and pasting the output. A full explanation of what you have done (see the section of the syllabus entitled Formatting and Grade Reconsideration Expectations) on those problems. If all you do is copy my SAS codes, you will have learned nothing and when exam time comes, you will pay the price. You will also lose formatting points.

Whereas type 1 problem solutions will never have to be submitted (these are for practice and to provide templates for SAS codes useful for solving type 2 problems and exam problems), **type 2 problem solutions will be collected one week after they are assigned.** These hand-in problems are mandatory problems for you (or your group, i.e., one paper per group) to submit. Cut and paste the complete solutions pages into a document file. Then each group (or individual student if you are not affiliated with a group) should email your submission to the TA, ShaoBo Li (lis6@mail.uc.edu). Generally, your submission must take place on or before 6:00 PM on the Thursday of the week following the assignment is given. With the resources available in the type 1 assignments, you should be well-prepared to understand the solutions for type 2 problems. The type 2 assignments will be collected and graded on the basis of 10 points each, 7 points for methodological correctness and completeness and 3 points for acceptable formatting. There will be a short duration for return (typically one class later). Submitted assignments will be subject to the group-work policy. That is, there will be one grade per group (or student if you are not affiliated with a group) and that grade will pertain to all members of the group. This implies that every group member must be concerned with quality control. Since SAS solutions are obviously available, I expect every paper will gain the full 8 points for methodological correctness and completeness. At the end of each module, the total point values will be adjusted to 100% and factored into your overall course grade as explain later in this syllabus. To summarize, to be successful in leaning statistical methods using state-of-the-art SAS software, use the type 1 problems and their solutions as a guide. Then attempt type 2 problems without looking at my solutions (where available). If you do this, then when the midterm problems are disseminated you will be well prepared to solve them quickly and correctly. If you do not, you may be sadly surprised to find that you do not have ready access to solution methods that will be necessary.

Except under exceptional circumstances, **as judged by me**, no late submitted assignments will be graded. Solving the problems is extremely important for reinforcing your comprehension of the material and in honing your computational skill with SAS software. If you do them faithfully, I guarantee that you will have gained a maximum amount of
knowledge in this course. You are welcome to discuss the solution of your problems with me or my graduate student during office hours or via email. But, I strongly urge you to collaborate with your colleagues in a responsible manner in attempting these problems, i.e., each party contributing equally to any cooperative arrangement. You are responsible for keeping current in the reading from the textbook and notes. Every submitted work will bear the names of the members of the work group (one submission per group). At the end of each module, an intra-group evaluation of the members of your group (excluding yourself) must be submitted. If you work alone no evaluation is needed. A copy of the intra-group evaluation form is in the Assignment Folder on the Blackboard site. Handing one in for each student is mandatory! Failure to do so will result in a possible grade reduction for the particular student. Since there are two sections of this course, 001 and 002, on all of your submitted work, please indicate which section you are in by including the appropriate section number on the front page of your submitted work near your name(s).

Grading: The final grade for the course will be a weighted average of your module 1 and 2 grades, with weights 2/3 for module 1 and 1/3 for module 2. The module 1 midterm exam will be worth 30% of the module 1 final grade. The take-home part of the module 1 final exam will be worth 30% of the module 1 final grade. The in-class part of the module 1 final exam will be worth 20% of the module 1 final grade. Exam scores will be adjusted so that a "100-90-80-70-etc." scheme for the corresponding letter grade equivalents "A-B-C-D-F" is appropriate. The collected hand-in problems that are graded will account for 10% of the module 1 final grade. The remaining 10% is under my discretion and will reflect my personal impressions based on sources which I feel reflect "attitude" such as class participation, cooperating with your work group in doing submitted assignments, helping your group learn SAS, intragroup evaluations, etc. For module 2, the break-up will be as follows: collected problems contribute 10%; take home part of module 2 final exam will be worth 40%; the in-class part of the mod 2 final will be worth 40%; the remaining 10% will be based upon other sources that reflect "attitude" such as class participation, cooperating with your work group in doing submitted assignments, helping your group learn SAS, intragroup evaluations etc. and my personal impressions. You will have the opportunity to give scores to your group members in an intra-group evaluation on two occasions. I will use these as part of my discretionary 10% (see section below).

Assurance of Learning – Learning Objectives: The grades on all assignments and test should comport with the learning objectives of this course. I will endeavor to adhere to these objectives in my teaching. I can only teach; you are the ones who must learn.

Upon completion of this course students will be able to:
1. Carry out statistical analyses on real world data by applying estimation, model fitting, and significance testing methodologies in an appropriate and correct manner.
2. Critically assess whether or not underlying assumptions for the use of the statistical methodology have been violated, and to take remedial measures if violation of these assumptions are detected.
3. Use high-level commercial statistical software, SAS for the statistical analyses on real world data.
4. To communicate the results of statistical analyses in language understandable to the general public such as a supervisor or colleague who may not have expertise in statistical methodology. See the section of the syllabus on formatting your work.

Students will have homework assignments and exams on which each of these objectives will be measured. For an overall measure of how well students are meeting these objectives, on every assignment and exam we will measure the percentage of students who score a “B” or better.

Group Work Structure of the Course: After the first class, each student may join a work group. A work group will consist of at least two students, but no more than six (I am flexible about exceptions). This work group should be maintained for the length of the semester. Under exceptional circumstances, the majority of a group can elect to ask a student to leave the group due to the student’s nonfeasance. Or a student may voluntarily withdraw. This should be a rare occurrence. The work group will cooperate in all work given during the semester including practice problems, studying, midterm and the take-home parts of the finals. The in-class portions of the final exams will require individual effort, and
is not to be a part of your group-work structure, although studying for it certainly is. All members of a group will share grades on any group submitted works. All members are to contribute equitably to the shared workload, carrying a fair weight for the burden. Group-work is not an invitation to slouch! At the end of the semester, members of each group will be asked to evaluate the overall contribution of the other work group peers on the basis of a number of criteria taking into consideration such factors as intellectual contribution, attendance at group meetings, mentoring and sharing knowledge, writing up the results, and running relevant SAS codes. The peer score will reflect, in some sense, an average over all of the work assigned as well as an average of the criterion above. Thus, a student in a work group who may have contributed much on one assignment, may not have contributed the majority of the work on another, yet still such work may be considered by other members to be meritorious “on the average”. This evaluation will be confidentially submitted to the TA, ShaoBo Li via email. A copy of the evaluation form is in the Blackboard site in one of the FINAL EXAM folders.

**EXCUSED ABSENCE POLICY:** I do not take attendance. While I do appreciate your courtesy in informing me of an absence, or if you must come to class late or leave early, it is not necessary to inform me. The faculty senate has issued a statement of policy regarding make-up work for days missed due to religious observance. It reads in part, “Any UC student who is unable to attend classes or participate in any examination, study or work requirement on some particular day(s) because of his or her religious belief should be given the opportunity either to make up the work that was missed or to do alternative work that is intrinsically no more difficult than the original exam or assignment — provided that the makeup work does not create an unreasonable burden upon University of Cincinnati and its faculty. Upon request and timely notice, students should be provided reasonable accommodation.” My personal policy is that if you need to miss a class or part of a class for religious reasons or any other reasons for that matter, it is acceptable to me. Take what you want from this experience. Since most exams are take-home affairs, the issue of excused absences is nullified anyway (with the exception of the in-class part of the final), and so I expect that all due dates will be honored. In addition, the information technology people tell me that they will record each class and make them available to you, so that a missed class can be reasonably recouped with some effort on your part. As stated elsewhere in this syllabus, you will have the opportunity to rate your team members’ participation and performance at the end of the each module. I will not offer any student any alternative work or make-up assignments.

**Miscellaneous (hodgepodge of things):** I subscribe to the UC policy on academic integrity and I will take appropriate action if I discover plagiarism or any student disregarding instructions on the limitations I place on consultation regarding exams/projects. ([http://www.uc.edu/conduct/Academic_Integrity.html](http://www.uc.edu/conduct/Academic_Integrity.html)). Please be aware of this code and abide by its provisions. In particular, cheating such as copying from another student’s paper will not be tolerated. A minimum of a failing grade on the exam or assignment in question will be imposed with possibly more serious consequences to follow. This has occurred in the past. In particular, do not consult any source or person outside of you own workgroup when producing exam/project material that you intend to submit for a grade. The grade of Incomplete will be given only for fully documented medical conditions or other catastrophes as judged solely by me. I strongly discourage a student going this route. No special exams/projects, or assignments will be given at semester's end for grade improvement. Students are responsible for all material, assignments and announcements made in class. All exams/projects will be carried out as described in the section on group work activity. The final exam will remain in my possession for a year. You should retain all other assignments or graded material throughout this course. If a notice of cancellation appears on the classroom door, or if you hear of such a cancellation, you must remain in the classroom for a period of ten minutes to confirm this action. I regard a student's attitude and interest to be a very important component in evaluation and I reserve the right to raise a grade based on my impression. I have a discretionary 10% of your grade in my possession. Please remember this. You are strongly urged to suggest any improvements in the teaching or classroom procedure. Every student will obtain a Blackboard account by going to [http://canopy.uc.edu/](http://canopy.uc.edu/) on the INTENET. We will use Blackboard to communicate. All of the class notes are on Blackboard. All of the SAS recipes I use in class are there too and much, much more. Please explore this site. You may drop the course without penalty up to the date set by UC for such eventualities (please look this up). I am often asked if a group can consist of students from different sections. This is a very bad idea for many reasons. But, I will permit it.
We will use SAS to solve some problems this semester (see the next section of the syllabus). You may wish to read the material in the SAS guide of general interest, i.e. the material on creating SAS data sets. Later on you will probably want to read the parts of later chapters which pertain to our class lectures and text. It is your responsibility to gain the proficiency necessary to enter, edit, submit and retrieve a SAS data set. SAS is available on computers in the LCB computer labs, as well as other places around campus. Using these machines carries the disadvantage of requiring your presence on campus. There is a way to contact the LCB computers remotely, so you do not need to be on campus. This entails downloading and running VMware View Client. I think you can get started by going to the URL, http://business.uc.edu/technology/services/ucvlabs.html. Another free way to access SAS is as follows: go to http://www.sas.com/en_us/software/university-edition.html. Download the SAS software and use it on any computer that you own. I do not like this alternative nor do I recommend it for most of you because the university edition of SAS does not have the suite of software called SAS GRAPH that enables you to create graphical displays similar to what I demonstrate in my SAS examples. In most other ways, however university edition will be fine and you may like it better than using UCvlabs. A third alternative that is not necessary if you take my first suggestion, is to buy a license for SAS from UCIT. I do not know the price exactly, but I think it is $71 (take your complaints to them). A few students in each class I teach with SAS as the analysis tool do obtain the license. I will supply all SAS recipes using SAS command line format. These can all be run under any version of SAS you use.

The Importance of SAS in this Course
We will be analyzing the data provided in the text problems and other problems. It is not only convenient to use a computer to aid in these analyses, it is vital that students develop a rudimentary proficiency in using a computer. Professionals and academics do not do the computations indicated in the formulas using paper and pencil any more. Therefore, I have chosen SAS as the preferred method to conduct the analyses. SAS is used in many companies as a way to manage and analyze data. This is a great opportunity to learn a few things about this powerful tool. I will provide you with many 'recipes' for implementing the analyses. Some of these are in the class notes. Others are in the SAS Files section in Blackboard. You should develop the ability to refer to the SAS on-line guides, or the Help function to search for guidance. Note that you do not have to buy any SAS manuals. All of them are available on SAS ONLINE Version 9.1, 9.2, 9.3, and 9.4: the URLs to visit are (there may be some new ones that you should try to find yourself):
http://support.sas.com/onlinedoc/913/docMainpage.jsp
or
http://support.sas.com/cdlsearch?ct=80000
All computer system inquiries may be addressed to the LCB computer consultants who should be well equipped in this regard. By the way, I’m using SAS 9.4 version, which is what is available using UCvlabs. Last year I used Version 9.3. Some of my SAS codes on Blackboard need to be modified to be usable in Versions 9.4 and higher.

Formatting and Grade Reconsideration Expectations:
The goal of this section of the syllabus is to help you develop good presentation habits. One of the major goals for this course is for students to learn how to clearly and concisely present technical results in the context of real world problems. As such homework problem solutions should read like a report you would give to your boss. The following guidelines are to help you practice this skill. On all assignments, 3 out of 10 points of the problem grade will be reserved for formatting. Therefore 7 points is reserved for correct mythological solution. Thus, a single homework problem grade may look like 7/1 which means that the solution (and code) was correct on that problems (7 points out of 7), but the formatting was deficient on the problem (1 point out of 3). The formatting points can have values: 0 = very bad or missing; 1 = deficient in a major way; 2 = good, but can be improved; 3 = very good. The total assignment grade will also have two parts, each part being the average solution grade.
and the average formatting grade computed separately. For example, say the two assignment average grades are 6.4 and 2.4, respectively. This means that the average of the solution grades is 6.4 out of 7 and the average of the formatting grades is 2.4 out of 3. The total assignment grade is $6.4 + 2.4 = 8.8$ out of 10. All three numbers are reported to you.

- In general most responses to the problems to the do not need to be more than one page. The TA will not grade you on this one but keep in mind that more is not always better when writing a report. An example of a correct and concise solution can be found later in this section of the syllabus. Don’t make things overly complicated.

- Please clearly label the questions and answers. Use context for all conclusions. SAS output or numbers with no labels or context will be counted as completely wrong and lead to a formatting grade deduction. A good rule of thumb is that the grader should be able to tell what the question is asking without referencing the material. Trust me your future boss is not going to look more than once at a confusingly organized or worded report.

- Please put all of your code and output (such as tables and figures) as the last item in the very back of each problem. The grader only really needs to see your code if your process seems correct but your numbers are wrong. Otherwise (just like your future bosses) the grader does not need (or want) to see your SAS code at all. You will want to reference a table or chart from the back of your assignment, so please number them and use that number as a reference. An example on the next page will give you a good idea of how to do this. If the question specifically requests a chart you may put it next to the question. Also if you are not using part of the SAS output then you should not include it in the report. It provides no benefit and is more likely to confuse the reader.

- Students always ask if they need to include formulas in their reports. The answer is, you must either provide a reference to a formula from the text book, notes, or alternatively report all mathematical formulas using mathematical symbols. For example, if you want to specify a sample standard deviation, then $s = \sqrt{\text{sum}((X_i-\text{Xbar})^2)/(n-1)} = 2.31$ should be written $s = \sqrt{\frac{\sum(X_i - \bar{X})^2}{n-1}} = 2.31$. I know this takes time but if you are going to be writing technical reports in the future you are going to need to get accustomed to this. Alternatively, to save time, just cite the text, e.g., “$s = 2.31$ (see formula 3.5.1 in Snedecor and Cochran)”. Usually, you can cite the place in the SAS output in which the value of $s$ can be found. Then you will need to cite the appropriate table at the back of the problem where this may be found. For example, “$s = 2.31$ (see Table 2)”. So you see, the tables can be useful and save you time.

- Please put the questions in order they are assigned. This is mostly for your benefit. If the questions are out of order the grader is more likely to miss that you completed a problem.

Finally, the TA has been instructed by me not to accept requests for more points on a problem unless there is a grading mistake. When the TA makes grading decisions they are implemented uniformly for everyone in the class. Change cannot be made on your grade without changing everyone else’s as well. Not only would this be a logistical nightmare but it is not a productive use of time. If a grading mistake has been made (which is bound to happen from time to time) or you really don’t understand what your mistake is please make an appointment with the TA. TA’s are happy to help try and explain any concepts you do not understand or grading decisions made. However before you come to the TA’s office please make sure you have checked the solutions and have identified a specific question to ask. I recommend writing any questions down to save time.

Example of good format in hypothesis testing problem: HW 7 #2.

THE PROBLEM: In a study of beer preferences and gender, members of a sample of 150 potential consumers were asked to express preference for each of three types of beers, Light, Regular, or Dark. It turned out that there were 80 males and 70 females in the sample. The data are in the following table.
Test the hypothesis of independence of gender and beer preference at the 0.05 significance level.

THE SOLUTION:

i) The hypotheses are:
   \( H_0: \) The two characteristics, customer gender and beer preference are independent
   \( H_1: \) The two characteristics, customer gender and beer preference are not independent

Perform the test at significance level \( \alpha = 0.05 \).

ii) The test statistic is the chi-square statistic and its value can be computed from the contingency table (see Table 1) as
   \[
   \chi^2 = \frac{\text{cells}(f_{ij} - \hat{F}_{ij})^2}{\hat{F}_{ij}}
   \]
   where \( f_{ij} \) is the observed frequency of cell \( ij \) and can be seen as the top number of each cell in Table 1. \( \hat{F}_{ij} \) is the expected frequency of cell \( ij \). And it can be seen as the second number in each cell of Table 1.

iii) Using the p-value approach, the decision rule is to reject the null hypothesis and conclude that the two characteristics are independent at the 0.05 level of significance if the p-value, namely \( P(\chi^2 > 6.1224) < 0.05 \).

iv) Conclusion: Since the p-value is 0.0468 (see p-value in SAS output Table 2 below), we can reject the null hypothesis. That is, conclude that the two characteristics, customer gender and beer preference are not independent at the 0.05 significance level.

Relevant output shows the following contingency table and chi-square information:

<table>
<thead>
<tr>
<th>Gender</th>
<th>Light</th>
<th>Regular</th>
<th>Dark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>20</td>
<td>40</td>
<td>20</td>
</tr>
<tr>
<td>Female</td>
<td>30</td>
<td>30</td>
<td>10</td>
</tr>
</tbody>
</table>

![Contingency Table](image)
Table 2

<table>
<thead>
<tr>
<th>Statistic</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>2</td>
<td>6.1224</td>
<td>0.0468</td>
</tr>
</tbody>
</table>

The SAS Code

/* Use contingency table approach with PROC FREQ*/
data pref;
input gender $ beer $ numcell @@;
cards;
male light 20 male regular 40 male dark 20
female light 30 female regular 30 female dark 10;
proc freq data=pref;
weight numcell;
tables gender*beer/ expected chisq; run;
/* Chi-Square test yields df= 2 Chisquare= 6.1224 p-value= 0.0468*/

Example of bad formatting for the same problem

=======================================================================
Solution:
=======================================================================
Example of good format for Problem 7 of HW 3

Problem: Compute a 95% confidence interval for the mean salary of all trained executives using the sample of 30 trained executives you selected in HW 1 using simple random sampling from the data set pay1.xls. Use the t-method.

SOLUTIONS: We can do this problem 3 ways:

First, the formula (Snedecor and Cochran 4.10.1) for a 95% CI for the mean of a normal population is \( \bar{X} \pm t_{29,0.025}S/\sqrt{n} \) where the sample mean, sample standard deviation of the sample of size n = 30, and the .975 quantile of the t\(_{29}\) distribution are,
\( \bar{X} = 52590.80 \), \( S = 4444.77 \), \( t_{29,0.025} = 2.0423 \), respectively. So the 95% CI for the mean salary is (50931.10, 54250.50). These are found in Table 1 using a SAS data step. Note the t-quantile was found using the SAS quantile function for this distribution.

Second, using PROC MEANS, the 95% confidence interval is given in Table 2 based on the option CLM with ALPHA =0.05.

Finally, using PROC UNIVARIATE with the option CIBASIC (TYPE=TWOSIDED ALPHA = .05), we get the result is Table 3.

The meaning of this 95% CI for the mean salary is: if random samples each of 30 executives were independently drawn repeatedly an infinite number of time, then 95% of the time the true population mean would lie within the limits \( \bar{X} \pm t_{29,0.025}S/\sqrt{n} \). Since the results on our sample of n = 30 gives (50931.1, 54250.5), we can say we have 95% confidence that the true mean salary lies within those limits.

Relevant Output and SAS Code:

| TABLE 1: Result of the Data Step analysis |
|---|---|---|
| t  | t_CI_lower | t_CI_upper |
| 2.04523 | 50931.10 | 54250.50 |

| TABLE 2: Result of PROC MEANS |
|---|---|---|---|---|
| Analysis Variable : Salary Salary |
| N | Mean | Std Dev | Lower 95% CL for Mean | Upper 95% CL for Mean |
| 30 | 52590.80 | 4444.77 | 50931.10 | 54250.51 |

| TABLE 3: Result of PROC UNIVARIATE |
|---|---|---|
| Basic Confidence Limits Assuming Normality |
| Parameter | Estimate | 95% Confidence Limits |
| Mean | 52591 | 50931 | 54251 |

The SAS code:

```sas
PROC IMPORT DATAFILE = "C:\pay1.xls" OUT = salaries replace dbms=xlsx;
   GETNAMES = yes;
RUN;
```
/* Sample of size 30 from HW 1*/
proc surveyselect data=salaries seed=45221
   sampsize=30 method=srs out=final;

PROC MEANS data=final N MEAN STD CLM ALPHA=.05;
   VAR salary; RUN;
/*Or using a SAS data step*/
data CI2;
   mean= 52590.80;STD= 4444.77;n=30;
t=quantile('t',.975,n-1);
t_CI_lower=mean-t*std/sqrt(n);
t_CI_upper=mean+t*std/sqrt(n);
proc print;var t t_CI_lower t_CI_upper; run;
/* Or use proc univariate*/
PROC UNIVARIATE data=final CIBASIC (TYPE=TWOSIDED ALPHA=.05);
   VAR salary; RUN;

---

Solution to Problem 7 HW3:

Homework and test reconsideration policy:  Due to the typical volume of questions regarding grades on assignments and tests that are handed back in class, I find that it is prudent to post a formal policy to keep things orderly and polite. Homework and tests should serve as gauges of your comprehension of the academic material in the course. In that

<table>
<thead>
<tr>
<th>Moments</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>30</td>
<td>Sum Weights</td>
<td>30</td>
</tr>
<tr>
<td>Mean</td>
<td>52590.8033</td>
<td>Sum Observations</td>
<td>1577724.1</td>
</tr>
<tr>
<td>Std Deviation</td>
<td>4444.77074</td>
<td>Variance</td>
<td>19755987</td>
</tr>
<tr>
<td>Skweness</td>
<td>-0.306283</td>
<td>Kurtosis</td>
<td>-0.3475375</td>
</tr>
<tr>
<td>Uncorrected SS</td>
<td>8.35457E10</td>
<td>Corrected SS</td>
<td>572923622</td>
</tr>
<tr>
<td>Coeff Variation</td>
<td>8.45181219</td>
<td>Std Error Mean</td>
<td>811.5004</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Basic Statistical Measures</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>Variability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>52590.80</td>
<td>Std Deviation</td>
<td>4445</td>
</tr>
<tr>
<td>Median</td>
<td>52320.00</td>
<td>Variance</td>
<td>19755987</td>
</tr>
<tr>
<td>Mode</td>
<td>.</td>
<td>Range</td>
<td>17751</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Interquartile Range</td>
<td>6856</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Basic Confidence Limits Assuming Normality</th>
<th></th>
<th>95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>Estimate</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>52591</td>
<td>50931</td>
</tr>
<tr>
<td>Std Deviation</td>
<td>4445</td>
<td>3540</td>
</tr>
<tr>
<td>Variance</td>
<td>19755987</td>
<td>12530511</td>
</tr>
</tbody>
</table>
sense, the grades are for you. But since professors must ultimately evaluate your performance, grades are also for the professor. Ideally, the grades should be accurate indicators of your mastery of the material, but of course, gauges are not always perfectly accurate since misunderstandings arise in the grading process. My philosophy about grading is expressed in the hope that the grading of your work is an aid to your education, and not a form of punishment. Below are my guidelines for reconsideration of grading on homework and tests.

1. In all discussions, especially the sensitive topic of your grade, we should always make every effort to be polite. This applies to all of us, including me, and my TA, ShaoBo Li, lis6@mail.uc.edu

2. Classroom time is not the time to request reconsideration of grading of your work. This time is reserved for instruction, demonstration of techniques, and answers to questions of general interest, rather than personal issues such as grades.

3. The first avenue for reconsideration should be my TA since he is the person closest to the source of the grade. Also, since he is the grader, he should know why points were deducted better than I. He is available in room 534 on a day and time to be announced in class. He may also help you with your computer issues.

4. If this time is not convenient to you, you can come to my office hours, which are posted on the syllabus. Or, you can make a special arrangement for a time outside of class that is mutually agreeable; you should do this via email or in person in my work space in room 106 since I will have my calendar in my office to consult whether the time you have chosen is feasible for me as well. I do not have my calendar in class.

5. In preparation for your visit, please write out your issues and concerns with as much detail as possible before hand. Then the discussion can flow more smoothly from these facts.

6. If you have not received satisfaction from my grader regarding your grading issues, you can make an appointment with me. I may resolve the issues or I may request a three-way meeting to settle any disputes.

7. One question that always comes up is “how long is the period in which I can ask for reconsideration?” It is strongly suggested that you review your homework paper within a week from when it is returned to you. In consideration for the teaching assistant, papers more than two week past the return time may not be reconsidered. That should give you plenty of time to get your issues resolved.

Further Remarks
This course is intended as a survey of statistical methodologies useful for research in a variety of disciplines which make use of data collected in experiments or observational studies and should appeal to students in business, economics, the social sciences, life sciences and engineering. Methods of estimation and significance testing using classical, nonparametric, and some robust techniques will be covered including an introduction to simple linear regression and ANOVA. Introduction to multiple regression via matrix methods will also be undertaken. The use of SAS for implementing data analyses on the computer is an important feature of the course as well. This course serves as a minimal prerequisite to our department's Statistical Modeling sequel courses 7042 offered in the Spring and Experimental Design (7026) which may or may not be offered in the Spring too. It should also serve as preparation for a variety of psychometric, sociometric, measurement and testing, and other statistical methods courses found throughout the university. A course outline, that will certainly be updated as time goes by, for module 1 and 2 follows.

MODULE 1
1. Introduction; Frequency Distributions
   -random sampling and examples
   -frequency distributions
   -probability distributions
2. The Mean and Standard Deviation; Normal Distribution
   - population and sample mean and standard deviation
   - coefficient of variation
   - normally distributed populations
   - distribution of the sample mean
   - the t-distribution
   - confidence interval estimation of the population mean
   - probability plots

3. Tests of Hypotheses
   - tests of the mean of a normal population
   - significance testing and confidence intervals
   - one and two tailed tests
   - power of a significance test
   - distribution of sample variance
   - interval estimates and tests for the population variance
   - goodness of fit tests: Chi square test
   - tests of skewness and kurtosis

4. Comparison of Two Populations
   - estimates and tests of differences
   - paired experiments
   - comparison of means from two independent samples
   - test of equality of two variances
   - paired vs. independent samples
   - sample size in comparative experiments

5. The Binomial Distribution
   - rules of probability
   - Bonferroni's inequality
   - normal approximation and continuity correction
   - significance test and confidence intervals for a proportion
   - comparison of proportions in paired samples
   - comparison of proportions in independent samples: 2x2 tables
   - sample size for comparing two proportions
   - the Poisson distribution

6. Nonparametric Methods
   - sample median
   - estimation of standard deviation from the sample range
   - sign test
   - signed-rank test
   - rank sum test for two independent samples
   - comparison of rank and normal tests
   - nonparametric confidence intervals
   - randomization tests

7. Analysis of Frequencies in One and Two-Way Classifications
   - single classification with more than two classes
-single classification with equal expectation
-test that Poisson samples have equal means
-two way classifications: 2xC contingency table
-test of homogeneity of binomial samples
-ordered classification
-test for linear trend in proportions
-RxC contingency tables
-sets of 2x2 tables

8. One-Way Classification: Analysis of Variance
-fixed effects analysis of variance
-the F test
-planned comparisons among class means: contrasts
-orthogonal comparisons
-unequal sample sizes
-weighted linear regression
-testing effects suggested by the data: Scheffé's method
-inspection of all differences between pairs of means: LSD

Module 2

9. Matrix concepts and multivariate probability distributions
-review of many matrix concepts such as transpose, inverse, determinants, eigenvalues, etc
-multivariate normal distribution
-moments of multivariate normal random vectors and matrices

10. Simple Linear Regression
-linear regression model
-analysis of variance for a linear regression
-method of least squares for estimating model parameters
-estimation and prediction of a response
-testing a suspected outlier
-linear calibration
-diagnosticss and remedial measure

11. Correlation
-sample correlation coefficient r and its properties
-bivariate normal distribution
-uses of the correlation coefficient: reliability measures
-testing for zero correlation
-confidence intervals and tests for correlations
-rank correlation
-matrix approach to correlation analysis

12. Multiple Linear Regression
-the model and parameter interpretation
-matrix approach to linear regression
-least squares estimation
-inference: estimation, prediction and hypothesis testing
-diagnosticss and remedial measures
-extra sum of squares principle
- principle of conditional error
- coefficient of partial determination
- standardized multiple regression