STATISTICS PROJECT REQUIREMENTS

Version 1.2
Milo Schield

INTRODUCTION:
This project can be very time consuming. And it may become very frustrating. The following is a proposed budget of your time. Plan your time carefully. What is your motivation to undertake this time consuming project? Suppose that your test scores do not seem to adequately demonstrate your knowledge of statistics. Perhaps this project will provide sufficient evidence for me to take the risk of concluding that you do know more about statistics than is shown in your test scores so that your grade should be significantly improved. A good project has saved many students from a near-failing grade.

TIME BUDGET: (assuming you know Minitab and have completed a sample analysis)
    1 hrs Selection of population, topic and variables (review with instructor)
    1 hrs Design of questionnaire and Sampling Technique (review with instructor)
    3 hrs Distributing survey and getting results [Watch your time on this one]
    2 hrs Minitab: Keyin data; check for errors; learn Minitab better.
    4 hrs Minitab: Part 3: Generate Data summaries with your comments!
    6 hrs Minitab: Part 4: Testing Hypothesis with your comments!
    3 hrs Executive summary; binding; checking, handwritten comments
    =====
    20 hours   Don't wait till the last minute!!!

COMMENTS: Time spent on questionnaire design, a non-convenience survey and data analysis pays off. Time spent keying lots of surveys with lots of variables doesn't pay!!!

GENERAL REPORTING REQUIREMENTS
Typed and single spaced (Last minute handwritten changes/comments OK)
Must use Minitab for data capture, data summary, data analysis and hypothesis testing.
    Use Minitab for sections 3 and 4; Scotch tape is OK; marked out errors are OK.

Reporting Requirements

1. Executive Summary (Title page plus 1 page) [Do this section last]
   1.0 Cover Sheet: Course, Title of Paper/Project, Your name and the date
       Your **<u>signature</u>** with "I created this survey, collected, entered and analyzed data,
       and prepared this report by myself with the following exceptions/assistance:..."
   1.1 Any last minute changes, recognition of errors, etc.
   1.2 Project Summary: Expected results and unexpected results
   1.3 Recommendations/comments on your particular project: future improvements,
etc.
   1.4 Recommendations/comments on overall project assignment.
           Estimated time spent and value received. Good/bad parts. Improvements..

2. Detailed Findings (General):
   2.1 Nature of the subject/population. Goal of survey
   2.2 Survey design and Sampling Technique: convenience, SRS, stratified, cluster, ...
           Technique: Number of surveys distributed, response rate, where/when/whom
           Possible bias or lack of preceision due to survey technique
   2.3 Survey questionnaire (copy with exact wording)
           Survey coding (can shown on questionnaire) Code Yes=1, no=0, no reply=*.
   2.4 Discuss sources of error (lack of precision and bias) for <u>each</u> variable.
           * lack of precision: Estimate whether this makes data unusuable.
           * bias: Estimate direction of bias {Estimate frequency and severity of bias}
                   If quantitative,<u> estimate the numerical magnitude</u> of bias on the mean
                   If qualitative, <u>estimate the magnitude</u> of the effect on the results.
   2.5Representative sample of target population: Discuss in relation to your survey.
           Any pre-survey or post-survey control to change composition of sample?

3. Data Summary (Generated by Minitab)
   3.1 Minitab printout of data (PRINT C1-C_). Name all columns. Printout of INFO.
   3.2 Minitab summary of each item of data (**<u>include your comments</u>**:
B 3.3 Minitab summary of some data from two populations
A 3.4 Minitab summary of some bivariate data
A 3.5 Minitab summary of multi-population breakout
       Note 3.2-3.5: See requirements on page 3

4. Detailed analysis for each hypothesis (Minitab generated; Requirements on page 3)
       Preferred: Select alternate hypothesis so that the data supports the alternative.
       # Written: <u>An exact statement of each Null and Alternate hypothesis</u>
       Minitab commands used to obtain relevant data (Copy C1 to C9;Use C2=1:3,5:8.)
       Minitab Dotplot & Describe (or Table) of data (even if presented elsewhere)
   Minitab statement of test and Minitab printout of the results.
       [Make sure your stated intention and your Minitab commands agree!!!!!!!]
       # Your conclusion(s) <u>with associated level of confidence</u>
       # Your analysis of results (How dependent is conclusion on level of confidence?)
   See page 4 for a detailed breakout of all parts of section 4.

Data Requirements

1. Obtain data on at least 40 subjects from a population of over 100 (no census)

2. Must include at least 10 variables (questions):
    at least 2 quantitative <u>continuous</u> variables with enough variation to facilitate a
        correlation or a regression (eg. Not the age of first-year students)
    at least 2 quantitative discrete variables (number of classes, number of siblings,etc.)
    at least 2 ordinal, 2 multi-nominal, and 2 binomial variables
        One ordinal must be a 5 point Lichert scale: 1=strongly dislike;5=strongly like
        Binomial data must be coded using 0 and 1 (to facilitate use of mean as fraction)

3.x Data Summary

3.2 UNIVARIATE DATA (one for each item of data from your survey)
    A. If binomial, calculate percents (MEAN). State the unit of measure. [p.42 top]
    B. If multinomial, ordinal or discrete, count responses (TALLY). [p.42 top]
    C. If quantitative or ordinal, summarize distribution (DESCRIBE) [p.43 bottom]
        Analyze effect of outliers: Are mean and trimmed mean signficantly different?
        [Extra credit: Calculate and comment on skewness and non-normality]
    D. If quantitative, plot data (DOTPLOT) [p.43 top]
        Comment on utility of mean as estimator (omit outlier, omit use of this data)

3.3 TWO POPULATION (Describe sample results comparatively)
    A. For quantitative data (by binomial): Show Minitab for at least three examples.
        (DESCRIBE data in Cc; BY index in Cb .[p. 00] Is TRMEAN diff.from MEAN?
         DOTPLOT of C_ ; BY index in C_.) [p.49 middle] Is difference due to outlier?
        Clearly state the sample conclusion (may be potential hypothesis about
population)
    B. For binomial data (by binomial). Show/summarize the results for at least one.
        DESCRIBE data in C_; BY index in C_.[p.00] Both columns must be binomial
        Wording: "% of Index:Value 1 with DATA > % of Index:Value 2 with DATA"

3.4 CORRELATION IN BIVARIATE DATA (Multiple attributes of one subject)
    1. Calculate correlation (CORR) for all quantitative and ordinal [p.56]
    2. Assume significant if absolute value > 2*SQRT(1/(n-1)) [Eg. 0.34 for n=37]
    3. Rank all significant correlations from positive to negative.
    4. Discuss any variations from what you might have expected.

3.5 MULTI-POPULATION ANALYSIS (Optional but necessary for an A) [p.44]
    Breakout mean of quantitative (and ordinal) data by two qualitative indexes
        (TABLE col index C_ row index C_; MEAN data in C_.)
    1. Show Minitab results
    2. State the most unusual/interesting comparison among the cells.
    3. [Add additional BY statements to obtain a finer breakout]

4. Hypothesis Tests

4.1 At least One test of independence/homogeneity for **each** of these Use Minitab
TABLE C_ by C_;CHISQ 1.) [p.45,46]. Comment/combine low count cells.
- A. Binomial by Binomial   Living (on/off campus) is independent of sex
- B. Ordinal by Binomial   Attitude to stats is independent of class (Fr., etc.)
- C. Multinomial by Binomial: Choice of major is independent of sex
- D. Ordinal vs Ordinal: Attitude to stats is independent of attitude to teacher
- E   Ordinal vs Multinominal   Attitude to stats is independent of major
- F. Multinominal vs multinominal   Choice of Major is independent of State

4.2 Quantitative data: At least **One** testable hypothesis for each of the following: [p.47]
1 pop: TTEST
- A. Use quantitative data   Height of average adult male > 66"
- B. Treat ordinal data as if it were quantitative   Average attitude to statistics > 2.5

4.3 Quantitative data: At least **One** testable hypothesis for each of the following: [p.49]
2 pop: TWOT (if stacked) or TWOS (if unstacked)
- A. Use quantitative data   Seniors are taller than Freshman
- B. Treat ordinal data as if it were quantitative   Seniors like math more than juniors

==

{Necessary for at least a B in addition to 4.1-4.3}
4.4 Binomial data:At least One testable hypothesis for each of the following: [p.52 top
1 pop: TTEST approximation (use short cut; treat like quantitiatve)
- A. Use existing binomial data   Less than 40% of seniors are male
- B. Map non-binomial data into binomialOver 60% of accounting seniors are female

4.5 Binomial data:At least One testable hypothesis for each of the following: [...........]
2 pop: TWOT or TWOS (use short-cut method; treat like quantitative)
- A. Use existing binomial data   More Seniors are males than Juniors
- B. Map non-binomial data into binomial   % of tall seniors > % of tall juniors

==

{Necessary for an A in addition to 4.1-4.5}   [see p. ......]
4.6 At least One pair of testable hypothesis (Beta_0 = 0 & Beta_1 = 0) for each of these:
- A. Regression of quantitative data vs quantitative (use pair with highest Corr)
- B. Regression of quantitative data vs ordinal (Indep.var) (Use pair with highest Corr)

Note: Comment on R^2. Assume significant if abs.value > 4/(n-1) (eg. 0.10 for
n=37)

Note: Summarize results in "English" (without reference to Beta 0 or Beta 1)

OPTIONAL ACTIVITIES

1. Use REGRESS;PREDICT & PLOT PLOT. On 2 regressions. Conf.Intrv vs Pred.Intrv
2. Analyze a sub-group of your sample (blue-eyed, blond, short males).[p.48]
3. Do multiple regression to improve explanatory power of your model.
4. Perform some independent, creative activity(ies) to impress your instructor
(What sample size would be required for a given margin of error on some variable?)

REPORT ANALYSIS FORM

REQUIREMENTS

	No	Little	Most	All	Exceed
Survey Requirements	0	1	2	3	*
1. At least 30 surveys					
2. Population of at least 100					

Data Requirements	<40%	60%	80%	100%	Exceed	*
2. 2 Quantitative (continuous, non-grouped)	__	__	__	__	__	
3. 4 ordinal (two lichert)	__	__	__	__	__	
4. 2 binomial	__	__	__	__	__	
5. 2 multinomial	__	__	__	__	__	

Formal Requirements:

6. Report Typed (handwritten 'adjustments' OK)	__	__	__	__	__
7. Followed reporting format	__	__	__	__	__

Survey Requirements

12. Bias on each variable(prob & nature)	__	__	__	__	__
13. Lack of precision on each variable	__	__	__	__	__

Conclusion:

Quant 1 pop	Wrong	Partly right	All right
Categorical	Wrong	Partly right	All right
Quant 2 pop	Wrong	Partly right	All right
Regression : Test on Beta1	Wrong	Partly right	All right
Regression: Prediction	Wrong	Partly right	All right

	1	2	3
OPTIONS:			
1. Survey had overall objective	No	Maybe	Yes
9. Type of survey	Convenience	SRS	Statified
10. Personal Initiative	None	Some	Lots
11. Alternate hypothesis could be True	Few	Most	All
12. State conclusions with Conf.Level	Seldom	Most	All
13. Use 'fail to reject Null' vs Accept Null	Seldom	Most	All

Name: _____ Topic: _____ Grade: _____

Coments: _____

QUALITATIVE REQUIREMENTS	Fail		Poor	Fair	Good	
Excel						
QUANTITATIVE REQUIREMENTS	No		Little	Most	All	Exceed
1 1 Typed;signed; changes; results; recomm.	0		1	2	3	4
2. 2.1 Discussed population & variables	0		1	2	3	4
3 2.2a Presented survey methodology	0		1	2	3	4
4 2.2b Used good survey methodology	0		1	2	3	4
5 2.3a&b Copy of survey with coding method	0		1	2	3	4
6 2.4 Discussed bias and lack of precision	0		1	2	3	4

----

7 3.1 Printout of raw data (Minitab PRINT)	0		1	2	3	4
8 At least 40 surveys (population >=100)	0		1	2	3	4
9 data: 4 quantitative (2 continuous; 2 discrete)	0		1	2	3	4
10 data: 2 binomial and 2 multinomial	0		1	2	3	4
11 data: 2 ordinal (one Lichert)	0		1	2	3	4

----

12 3.2a&bBinom (mean) & Ord/Multi (Tally)	0		1	2	3	4
13 3.2c&dOrd&Quant: Describe & Dotplot	0		1	2	3	4
14 3.3a 2 Population: Quant. by binomial	0		1	2	3	4
15 3.3b 2 pop: Binomial (Use proper wording)	0		1	2	3	4
16 3.4 CORR (All quant & ordinal)	0		1	2	3	4
17 3.5 TABLE Cb Cb;Mean Cq		0	1	2	3	4

----

18 4.1 CHISQ: Show/calc. cutoff	0		1	2	3	4
19 4.2a TTEST: 1 pop Quantitative	0		1	2	3	4
20 4.2b " " " " " : " " " Ordinal	0		1	2	3	4
21 4.3a TWOT: 2 pop Quantitative (or TWOS)	0		1	2	3	4
22 4.3b " " " " : : : : : : Ordinal	0		1	2	3	4

===

23 4.4a TTEST: 1 pop: Binomial		0	1	2	3	4
24 4.4b " " " " " " " " : Map non binomial	0		1	2	3	4
25 4.5a TWOT: 2 pop: Binomial (or TWOS)	0		1	2	3	4
26 4.5b " " " " " " " " : Map non binomial	0		1	2	3	4

===

27 4.6a Regress: Quant vs Quant: B0 and B1	0		1	2	3	4
28 4.6b " " " " " : Quant vs Ordinal: B0 and B1	0		1	2	3	4

OPTIONAL---

29 Regression: Predictionwith PLOT PLOT	0		1	2	3	4
30 Regression: Multiple	0		1	2	3	4
31 Analyze some sub-group	0		1	2	3	4
32 Analyze of assumptions (normality, etc.)	0		1	2	3	4
33. Other personal initiative	0		1	2	3	4