

Case-based Social Statistics II

CSSS 322

Spring 2002

Solutions to Practice First Examination

Exam: Tuesday, May 7, 10:30am - 12:00pm

Professor: Mark S. Handcock

Name: Ronald Aylmer Fisher

1. Please write your name in the above space.
2. **You need to do all 4 questions.** All questions are of equal value (but not necessarily of equal difficulty). attempted.
3. Do not turn the page until so instructed. (You will have 90 minutes to work after the examination has been discussed with you.)
4. You may use your crib sheet. Otherwise this is a closed book examination.
5. If you do not have enough room for your work in the place provided, use the back of a nearby page. (However, be sure to mark clearly which problem the material on the back of any page refers to.) If you pull the pages apart, sign all pages.
6. Answers should unambiguously state, in words, the approach taken. You should show your work so that partial credit can be given. Poorly described solutions will be penalized. unsupported answers
7. Good luck!

Question	Subject	Points available	Points earned
1	Incoming Notebooks	25	
2	A Bond Barometer	25	
3	Vitamin C	25	
3	Diversified Stock Funds	25	
Total		100	

Question 1) Incoming Notebooks (25 points)

The President of the University of Washington is concerned that the current students will have different views than the incoming students. He decides to survey the incoming class to see if they are in favor of requiring notebooks.

The pool of potential incoming students is quite large (many thousands) and so to reduce costs he conducts a survey of 100 randomly chosen incoming students. Of these, 60 indicate that they are willing to buy a notebook.

If more than 50% of the incoming class of students are willing to buy a notebook he will require it of all students. Research has shown that exactly 50% of past students were willing to buy a notebook.

Your objective is to formulate and execute a statistical hypothesis test for this situation. An error rate of 0.01 ($\alpha = 0.01$) should be used.

a) (5 points)

Give the null and alternative hypotheses in symbols and words.
Be sure to identify **all** symbols that you use.

Solution: The null hypothesis is

$$H_0 : \pi = 0.5$$

against the alternative hypothesis

$$H_1 : \pi > 0.5$$

Here π is the proportion of all incoming students who are willing to buy a notebook. This is a one-sided alternative. A two-sided alternative is not acceptable, i.e.,

$$H_1 : \pi \neq 0.5$$

The null hypothesis states that the proportion of all incoming students who are willing to buy a notebook is 0.5. The alternative hypothesis states that the proportion of all incoming students who are willing to buy a notebook is greater than 0.5. Clearly the latter is what we want to prove and the former is the “established” position based on experience with past students.

b) (5 points)

Construct a confidence interval appropriate for this test.

Solution: The point estimate of π is $f = 60/100 = 0.6$. The standard error of the estimate is $S_f = \sqrt{f(1-f)/\sqrt{n}} = \sqrt{0.6(1-0.6)/\sqrt{100}} = 4.9\%$. The confidence interval (for the one-sided test) is then:

$$f \pm t_{\alpha}(n-1) \times s_f$$

Here $\alpha = 0.01$ so the interval is:

$$\begin{aligned} &0.6 \pm 2.326 \times 0.049 \\ &0.6 \pm 0.114 \\ &=(48.6\%, 71.4\%) \end{aligned}$$

Note that as $n > 40$, we approximate the t -multiplier by that with infinite degrees of freedom, 2.326.

c) (3 points)

Briefly describe the meaning of Type I error in this case?

Solution: A Type I error is to decide that the null hypothesis is false when in fact it is true. In this case that means to decide that the proportion in the incoming class is above 50% when in fact it is 50%.

d) (4 points)

Briefly describe the meaning of Type II error in this case?

Solution: A Type II error is to decide that the null hypothesis is true when in fact it is false. In this case that means to decide that the proportion in the incoming class is 50% when in fact it is above 50%.

e) (4 points)

Given the above confidence interval, is the probability of Type II error large or small. Briefly say why.

Solution: Our best estimate of π is $f = 0.6$. If this were the actual value then we will not reject the null hypothesis if the confidence interval includes 50%. From b), 50% is in the edge of the interval, so it will happen quite often.

f) (4 points)

The 98%, 95% and 90% two-sided confidence intervals for the proportion in the entire incoming class are:

98% (48.6, 71.4)

95% (50.4, 69.4)

90% (52.0, 68.0)

Based on these and b), what is the p -value for the hypothesis in a)?

Solution: At the α error level, we will reject the null if 0.5 is outside the confidence interval for π . The one-sided intervals can be derived by including values greater than the upper bound (i.e., in the direction indicated by the alternative hypothesis). The intervals are then:

99% > 48.6

97.5% > 50.4

80% > 52.0

The 97.5% confidence interval does not contain 50%, but the 99% interval does. Thus the largest value of α is between 0.025 (from d) and 0.01. This latter interval is almost exactly right, so we expect it to be closer to $p = 0.025$.

Question 2) A Bond Barometer (25 points)

As a guide for investment the *Wall Street Journal* publishes information on the total returns for various categories of bonds. The data below is from the last exam. It is from *The Wall Street Journal*, 10/8/93, R2. It compares a prediction of the total return for the next twelve months (based on the last quarter) to return over the previous twelve months for 17 categories of bonds.

Assume that these bonds are a random sample from the population of all bonds available.

<i>Fund Category</i>	Total Return	
	<i>Previous Year</i>	<i>Prediction for the next Year</i>
Zero Coupon Bonds	24.9	28.8
Convertible Bonds	21.3	20.4
Long-Term U.S. Treas.	15.1	18.1
Global Bonds	11.6	17.3
General Muni. Bonds	13.1	13.4
Flexible Income Funds	11.8	12.6
Longer-Term U.S. Govt.	10.0	10.8
Intermediate Corporate	9.5	10.8
Intermediate U.S. Treas.	9.3	10.2
Intermediate U.S. Govt.	8.0	8.9
High-Yield Junk	15.4	7.8
Short-Term U.S. Treas.	5.6	6.6
Short-Term Corporate	5.9	6.4
Short-Term U.S. Govt.	5.6	6.2
GNMA	6.7	4.3
Adjust. Rate Mort.	4.1	3.9
Short-Term Global Inc.	3.8	2.2
mean	10.69	11.10
standard deviation	5.90	6.87

a) (6 points)

Construct a 95% confidence interval for the mean total return for the previous year.

Solution: A 95% confidence interval for the total return, μ_{1992} , of the population of bonds for the previous year is

$$\begin{aligned}
 & \left(\bar{X}_{1992} - t_{\alpha/2}(n-1) \times \frac{s_{1992}}{\sqrt{n}}, \bar{X}_{1992} + t_{\alpha/2}(n-1) \times \frac{s_{1992}}{\sqrt{n}} \right) \\
 & = \left(\bar{X}_{1992} - t_{0.025}(16) \times \frac{s_{1992}}{\sqrt{n}}, \bar{X}_{1992} + t_{0.025}(16) \times \frac{s_{1992}}{\sqrt{n}} \right) \\
 & = \left(10.69 - 2.120 \times \frac{5.90}{\sqrt{17}}, 10.69 + 2.120 \times \frac{5.90}{\sqrt{17}} \right) \\
 & = (7.66\%, 13.72\%)
 \end{aligned}$$

as $\bar{X}_{1992} = 10.69$, $n = 17$, $s_{1992} = 5.90$, $\alpha = 0.05$.

b) (7 points)

Construct a 99% confidence interval for the mean total return for the prediction of the next year.

Solution: A 99% confidence interval for the predicted total return, μ_{1993} , of the population of bonds for the next year is

$$\begin{aligned} & \left(\bar{X}_{1993} - t_{\alpha/2}(n-1) \times \frac{s_{1993}}{\sqrt{n}}, \bar{X}_{1993} + t_{\alpha/2}(n-1) \times \frac{s_{1993}}{\sqrt{n}} \right) \\ & = \left(\bar{X}_{1993} - t_{0.005}(16) \times \frac{s_{1993}}{\sqrt{n}}, \bar{X}_{1993} + t_{0.005}(16) \times \frac{s_{1993}}{\sqrt{n}} \right) \\ & = \left(11.10 - 2.921 \times \frac{6.87}{\sqrt{17}}, 11.10 + 2.921 \times \frac{6.87}{\sqrt{17}} \right) \\ & = (6.23\%, 15.97\%) \end{aligned}$$

as $\bar{X}_{1993} = 11.10$, $n = 17$, $s_{1993} = 6.87$, $\alpha = 0.01$.

c) (5 points)

Based on the results of a) and b), do you think the mean total return for the previous year and the mean prediction of the total return for the next year are identical? Briefly say why.

Solution: The claim is that predicted total return, μ_{1993} , of the population of bonds for the next year is equal to the total return, μ_{1992} , of the population of bonds for the previous year. In symbols this is

$$H : \mu_{1992} = \mu_{1993}$$

If they are equal then their confidence intervals should have substantial overlap. We see that the confidence intervals are almost identical (noting that one is 95% and the other is 99%). As such, there is little evidence in the data that the mean total return for the previous year and the mean prediction of the total return for the next year are different. That is, we do not reject the hypothesis.

d) (7 points)

One Wall Street pundit claims that the mean total return for the bonds for the next year will be over 12%.

Translate this claim into a hypothesis about the mean total return for the bonds for the next year. Express the hypothesis in symbols.

Based on your results in part b), do you think this pundit's claim is consistent with the data on the predictions for next year?

Solution: Expressing, as above, the predicted total return of the population of bonds for the next year by μ_{1993} , the hypothesis is that this value is over 12%. In symbols this is

$$H : \mu_{1993} > 12\%$$

As the above 99% confidence interval for μ_{1993} includes 12%, there is little evidence in the data that the mean total return for the bonds for the next year is not 12% as claimed. Note that even though 12% falls in the upper range of the interval estimate, this is not evidence, in and of itself, that the population mean is not 12%.

Question 3) Vitamin C (25 points)

To determine how well vitamin C prevents colds, a study in Toronto under T. W. Anderson divided volunteers randomly into two groups of 400 each - a treatment group given vitamin C, and a control group given a placebo (a 'dummy' pill that had no drug, but looked indistinguishable from the vitamin C pill). The proportion of people who caught colds during the winter was 74% for the treatment group and 82% for the control group.

In answering the following questions, suppose that the volunteers represent random samples from the entire Canadian population.

a) (3 points)

What was the purpose of the placebo?

Solution: The purpose of the placebo is to remove the effect of "taking a pill" from the study. That is, the volunteer getting better just because he or she is receiving **some** treatment. Our objective is to measure the effect of vitamin C, rather than other potential effects.

b) (5 points)

Construct a 95% confidence interval for the proportion of colds in Canada if everyone took vitamin C.

Solution: Let X be the random variable describing the number of people out of 400 who would catch cold during the winter while taking vitamin C. Then X is binomial with probability of success P_X and $n = 400$ trials. Let $\bar{P}_X = 74\%$ be the sample proportion of people who catch colds during the winter. As $n > 30$, we can use the normal approximation to the binomial proportion to calculate the confidence interval:

$$\left(\bar{P}_X - z_{\alpha/2} \times \sqrt{\frac{\bar{P}_X(1 - \bar{P}_X)}{n}}, \bar{P}_X + z_{\alpha/2} \times \sqrt{\frac{\bar{P}_X(1 - \bar{P}_X)}{n}} \right)$$

Here $\alpha = 0.05$ so the interval is:

$$\begin{aligned} & \left(0.74 - 1.96 \times \sqrt{\frac{0.74 \times 0.26}{400}}, 0.74 + 1.96 \times \sqrt{\frac{0.74 \times 0.26}{400}} \right) \\ & = (69.7\%, 78.3\%) \end{aligned}$$

Hence a 95% confidence interval for the proportion of vitamin C takers suffering from colds is

$$(69.7\%, 78.3\%)$$

c) (5 points)

Construct a 95% confidence interval for the proportion of people who would catch cold in Canada if everyone took the placebo only.

Solution: Let Y be the random variable describing the number of people out of 400 who would catch cold during the winter while taking a placebo. Then Y is binomial with probability of success P_Y and $m = 400$ trials. Let $\bar{P}_Y = 82\%$ be the sample proportion of people who catch colds during the winter. As $m > 30$, we can use the normal approximation to the binomial proportion to calculate the confidence interval:

$$\left(\bar{P}_Y - z_{\alpha/2} \times \sqrt{\frac{\bar{P}_Y(1 - \bar{P}_Y)}{m}}, \bar{P}_Y + z_{\alpha/2} \times \sqrt{\frac{\bar{P}_Y(1 - \bar{P}_Y)}{m}} \right)$$

Here $\alpha = 0.05$ so the interval is:

$$\begin{aligned} & \left(0.82 - 1.96 \times \sqrt{\frac{0.82 \times 0.18}{400}}, 0.82 + 1.96 \times \sqrt{\frac{0.82 \times 0.18}{400}} \right) \\ & = (78.2\%, 85.8\%) \end{aligned}$$

Hence a 95% confidence interval for the proportion of placebo takers suffering from colds is

$$(78.2\%, 85.8\%)$$

d) (4 points)

A 95% confidence interval for the decrease in proportion of colds in Canada that would be caused by everybody taking vitamin C is

$$(2.3\%, 13.7\%)$$

Based on this information and the previous results, write a brief verbal conclusion about the effect of vitamin C.

Solution: The effect of vitamin C on the proportion of the Canadian population catching colds over winters is, with 95% confidence, between 2% and 14%. An estimate of the mean effect of vitamin C is 8%.

The overall conclusion is that vitamin C has a modest, but positive effect in reducing the possibility of colds.

Question 4) Diversified Stock Funds (38 points)

Shares in diversified stock funds are traded among investors at market prices that need not equal the fund's net asset value. Below are the percent difference between the net asset value and the market price, that is,

$$\frac{100\% \times (\text{net asset value} - \text{market price})}{\text{net asset value}}$$

for 15 diversified stocks funds reported in the *Wall Street Journal* in March 9, 1990. The right-most column are the percent difference for the same stocks two days ago (Monday, October 12, 1992).

<i>Diversified Stock Fund</i>	Percent Difference	
	<i>March 9, 1990</i>	<i>October 12, 1992</i>
Adams Express	-9.7	-2.2
Baker Fentress	-16.1	-14.4
Blue Chip Value	-9.6	+7.6
Clemente Global	-18.4	-10.6
General American	-16.6	+1.3
Jundt Growth Fund	-5.5	-10.8
Liberty All-Star Equity	-14.0	-18.0
Quest For Value Capital	-15.9	-1.9
Quest For Value Income	-12.2	+34.5
Royce Value Trust	-10.6	-4.5
Salomon Fund	-7.4	-4.2
Source Capital	+0.2	+9.3
Tri-Continental	-14.1	-8.3
Worldwide Value	-5.8	-6.6
Zweig Fund	+6.2	+17.2
mean	-9.97	-0.76
standard deviation	6.74	13.46

a) (5 points)

One financial theory of asset pricing claims that the net asset value and market price of the diversified stock funds should, on average, be equal.

Translate this claim to a null and alternative hypothesis expressed in symbols. Be sure to define the symbols you use.

Solution: Let μ be the mean percent difference between the net asset value and the market price for the population of stock funds. The claim is that this mean is zero. Formally,

$$H_0 : \mu = 0$$

$$H_1 : \mu \neq 0$$

Note that the alternative hypothesis is two sided.

b) (6 points)

Test the hypothesis in a) based on the sample of stock funds in March 9, 1990. Use the 99% level of confidence. What is your conclusion.?

Solution: The test statistic is

$$t = \frac{\bar{X} - 0.0}{s/\sqrt{n}} = \frac{-9.97 - 0}{6.74/\sqrt{15}} = -5.73$$

As $|t| = 5.73 > 2.977 = t_{0.005}(14)$, we reject the null hypothesis that the mean percent difference between the net asset value and the market price for the population of stock funds is the same, at the 99% level of confidence.

c) (4 points)

Test the same hypothesis at the same level of confidence based on the sample of stock funds in October 12, 1992. What is your conclusion.?

Solution: The test statistic is

$$t = \frac{\bar{Y} - 0.0}{s/\sqrt{n}} = \frac{-0.76 - 0}{13.46/\sqrt{15}} = -0.22$$

As $|t| = 0.22 < 2.977 = t_{0.005}(14)$, we do not reject the null hypothesis that the mean percent difference between the net asset value and the market price for the population of stock funds is the same, at the 99% level of confidence.

d) (10 points)

The change in the mean percent difference between the net asset value and the market price for the population of stock funds over time is of great financial interest. Use the two samples to provide an 95% confidence interval for the difference in the mean percent difference between *March 9, 1990* and *October 12, 1992*. (You may assume that the distribution of the samples differ only in location if they differ at all).

Solution: Let X , Y represent the mean percent difference between *March 9, 1990* and *October 12, 1992*, respectively. The pooled sample variance is

$$\begin{aligned} s_p^2 &= \frac{(m-1)s_X^2 + (n-1)s_Y^2}{(m+n-2)} \\ &= \frac{(15-1)6.74^2 + (15-1)13.46^2}{(15+15-2)} \\ &= 113.3 \end{aligned}$$

so $s_p = 10.64\%$. As $m+n-2 = 28 < 30$, we will use the t multiplier, $t_{\alpha/2}(28)$. A 95% confidence interval for the change in percent difference is

$$\begin{aligned} &\bar{X} - \bar{Y} \pm t_{0.025}(28)s_p\sqrt{\frac{1}{m} + \frac{1}{n}} \\ &= -9.97 - (-0.76) \pm 2.048 \times 10.64 \times \sqrt{\frac{1}{15} + \frac{1}{15}} \\ &= (-17.17\%, -1.25\%) \end{aligned}$$

As zero is above the confidence interval, we can state that, with at least 95% confidence, that the mean percent difference of *March 9, 1990* is greater than the mean percent difference of *October 12, 1992*.

e) (10 points)

The same stock funds were used to measure the net asset value and the market price. Used the paired method to construct a 95% confidence interval for the mean percent difference between the net asset value and the market price for the population of stock funds. The sample standard deviation of the differences is 12.46%.

Solution: Let X , Y represent the mean percent difference between *March 9, 1990* and *October 12, 1992*, respectively. Then denote the differences by $D = X - Y$. We are told that the sample standard deviation of the differences is $s_D = 12.46\%$ and the mean difference is $\bar{D} = -9.97 - (-0.76) = -9.21\%$. As $n - 1 = 14 < 30$, we will use the t multiplier, $t_{\alpha/2}(28)$. A 95% confidence interval for the change in percent difference is

$$\begin{aligned} & \bar{D} \pm t_{0.025}(14) s_D \sqrt{\frac{1}{n}} \\ & = -9.21 \pm 2.14 \times 12.46 \times \sqrt{\frac{1}{14}} \\ & = (-16.33\%, -2.08\%) \end{aligned}$$

As zero is above the confidence interval, we can state that, with at least 95% confidence, that the mean percent difference of *March 9, 1990* is greater than the mean percent difference of *October 12, 1992*.

f) (5 points)

Which of the above two intervals is more appropriate (the independent or the paired)?

Solution: As the same stock funds are used for both measurements, there is reason to believe that there is dependence between the measurements. Hence the paired method is more appropriate.