Exam 1: MATH 4773/5773 FALL 2019

Dr. Wayne Stewart

Oct 4, 2019

Abstract

Please answer any 3 of the 4 questions – there are 5 sub-parts in each question, each worth 2 points.

Contents

1 Question 1

A	B	$\cdots$	K
n_1	n_2	$\cdots$	n_k

where $A, B, \cdots, K$ are categories and n_i are frequencies such that $n = \sum_i n_i$.

Please answer the following using the CLT and any other relevant results.

1.1 a)

If $\hat{p}_i = \frac{n_i}{n}$ find:

1. $E(\hat{p}_i)$
2. $E(\hat{p}_i - \hat{p}_j)$

1.2 b)

Derive the formula for $V(\hat{p}_i)$

1.3 c)

Derive the expression for $V(\hat{p}_i - \hat{p}_j)$ by showing all intermediate steps and explaining each step. The final expression must be in terms of n and $\hat{p}^{'}s$. You may use the result that $Cov(\hat{p}_i, \hat{p}_j) = \frac{-p_i p_j}{n}$

1.4 d)

Using the above, prove the following result:

A large sample $(1-\alpha)100\%$ confidence interval for $p_i - p_j$ in a one way table is

$$(\hat{p}_i - \hat{p}_j) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_i(1-\hat{p}_i) + \hat{p}_j(1-\hat{p}_j) + 2\hat{p}_i\hat{p}_j}{n}}$$

You MUST give explanations for each step and without explanation there will be NO points!

1.5 e)

Answer the questions following the description below:

```
irr=myread("irrigate.xls")
head(irr)
```

```
## # A tibble: 5 x 2
##   STRATEGY  FREQ
##   <chr>    <dbl>
## 1 A           17
## 2 B           27
## 3 C           22
## 4 D           15
## 5 E           19
```

9.8 *Irrigating cropland.* Because of erratic rainfall patterns and low water-holding capacities of soils in Florida, supplemental irrigation is required for producing most crops. A research team has developed five alternative water-management strategies for irrigating cropland in central Florida. A random sample of 100 agricultural engineers was interviewed and asked which of the strategies he or she believes would yield maximum productivity. A summary of their responses is shown in the table.

IRRIGATE

Strategy	A	B	C	D	E
Frequency	17	27	22	15	19

Figure 1:

```
irv=with(irr, rep(STRATEGY,FREQ))
tab=table(irv)
tab
```

```
## irv
##  A  B  C  D  E
## 17 27 22 15 19
```

a) Find the Pvalue for testing $H_0 : p_1 = p_2 = p_3 = p_4 = p_5 = 1/5$
b) What do you conclude concerning the NULL hypothesis?
c) Find all pairwise cis, $p_i - p_j$ with 88% confidence adjusted for multiple comparisons.
d) Using the pairwise cis generated above find the confidence interval for $p_2 - p_4$

2 Question 2

The following questions relate mostly to projection matrices and the notation is that which we have discussed in class:

2.1 a)

After drawing the apropriate picture, derive the expression for H.

2.2 b)

Show that $HX = X$ where $X =$ The Design Matrix

2.3 c)

Show that $I - H$ is idempotent.

Suppose that H_w projects Y onto the subspace spanned by the columns of the reduced design matrix X_r and H_Ω projects Y onto the space spanned by the columns of the complete design matrix X_c.

Show $H_\Omega - H_w$ is idempotent – you must justify any statement made!

2.4 d)

Prove that

$$SSE = \epsilon^{'}(I - H)\epsilon$$

2.5 e)

If $TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2$ Show that

$$TSS = Y^{'}(I - \frac{1}{n}J_n)Y$$

where $J_n = (1)_{n \times n}$

3 Question 3

The following has to do with expressions for estimators and covariance logic.

3.1 a)

Using $cov(CY) = Ccov(Y)C^{'}$ prove that

$$cov(\hat{\beta}) = \sigma^2(X^{'}X)^{-1}$$

3.2 b)

Suppose we have an **SLR** model

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

Find the 2 by 2 matrix expression for the $cov(\hat{\beta})$

3.3 c)

What is the expression for $Var(\hat{\beta}_0)$

3.4 d)

If $\bar{x} = 0$ what is the numerical value of $cov(\hat{\beta}_0, \hat{\beta}_1)$?

3.5 e)

What is the algebraic expression for $\sigma^2_{\hat{\beta}_1}$

Example 12.5

Cost Model with a Qualitative Independent Variable

A large consulting firm markets a computerized system for monitoring road construction bids to various state departments of transportation. Since the high cost of maintaining the system is partially absorbed by the firm, the firm wants to compare the mean annual maintenance costs accrued by system users in three different states: Kansas, Kentucky, and Texas. A sample of 10 users is selected from each state installation and the maintenance cost accrued by each is recorded, as shown in Table 12.6.

a. Do the data provide sufficient evidence (at $\alpha = .05$) to indicate that the mean annual maintenance costs accrued by system users differ for the three state installations?

b. Find and interpret a 95% confidence interval for the difference between the mean costs in Texas and Kansas.

Solution

a. The model relating $E(y)$ to the single qualitative variable, state installation, is

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

where

$$x_1 = \begin{cases} 1 & \text{if Kentucky} \\ 0 & \text{if not} \end{cases} \qquad x_2 = \begin{cases} 1 & \text{if Texas} \\ 0 & \text{if not} \end{cases}$$

BIDMAINT

TABLE 12.6 **Annual Maintenance Costs**

	State Installation		
	1: Kansas	2: Kentucky	3: Texas
	$ 198	$ 563	$ 385
	126	314	693
	443	483	266
	570	144	586
	286	585	178
	184	377	773
	105	264	308
	216	185	430
	465	330	644
	203	354	515
Totals	$2,796	$3,599	$4,778

Figure 2:

4 Question 4

The following question comes from Example 12.5 in the text MS 6th edition. You can use the s20x library to calculate interval estimates ciReg() etc. Use my transformation of the variables to ensure consistency. There are parts a) and b) from the text as well as c) d) and e) supplied by me. These will need to be answered completely.

4.1 a)

See above

4.2 b)

See above

4.3 c)

Is the model adequate?

4.4 d)

What is the estimate of the model standard deviation?

4.5 e)

What is the value of the χ^2 statistic that will produce the rejection region for the F-test on the last line of the summary(ylm) output?

```
bid=myread("BIDMAINT.xls")
#head(bid)
bid = within(bid,{
  y<-COST
  x1<-as.integer(X1)
  x2<-as.integer(X2)
})
```