Quiz Practice Questions 4 (Attendance 8) for Statistics 503
Introduction to Statistics
Material Covered: Chapter 6 Rao and Kuhn

These are practice questions for the quiz. The quiz (not the practice questions) is worth 5% and marked out of 5 points. One or more questions is closely, but not necessarily exactly, related to one or more of these questions will appear on the quiz. These practice questions are *not* to be handed in. Quizzes are to be done *using Vista* on the Internet **before** 4am of the date of the quiz. Vista will *not* allow any quiz to be done late. It is *highly* recommended that you complete this practice quiz, by hand, *before* logging onto Vista. The quiz is an **individual** one which means that each student does this quiz by themselves without help from others. Also check out *previous* quizzes given at

http://www.pnc.edu/faculty/jkuhn/courses/previous/quizzes/quizzes.html

1. Statistical Research Methods in the Life Sciences (Rao) Questions.

Section	Exercise(s)	hints
6.1, pages 205–206	(6.1)	see below
6.2, pages 219–222	(6.3)	see below
	(6.5)	see below
	(6.9)	see below
6.3, pages 239–241	(6.13)	look below
	(6.17)	look below
	(6.19)	look below
Extra Exercises	(1)–(3)	look below
Summary of quiz material		look below

Extra Exercises: Flu Symptoms and Aspirin, One Proportion Specified. The *observed* data from a random sample of 354 students from PU/NC in an investigation of the effect of aspirin on reducing flu symptoms is given in the table below.

observed, f	aspirin	no aspirin	subtotals
flu symptoms reduced	120	81	201
flu symptoms not reduced	57	96	153
subtotals	177	177	354

We know (or, at least, we assume we know) that 50% of *all* students at PU/NC take aspirin when they have the flu (and so 50% do not take aspirin when they have the flu). Does the observed data from the random sample of these 354 students contradict what we expect to see if whether or not flu symptoms are reduced is *independent* of whether or not aspirin is taken at $\alpha = 0.01$?

1. *Preliminary Analysis: Figuring Out The Test Statistic.*

   (a) The *observed* number of students where flu symptoms are reduced when aspirin is administered, is (circle one) **50 / 120 / 153**.

   (b) The *observed* number of students who took aspirin is (circle one) **50 / 150 / 177**.

   (c) The *expected* number of students, of 354, to take aspirin is $354 \times 0.50 =$ (circle one) **50 / 150 / 177**.
   Notice, in this sample, that since we assumed 50% took aspirin, the observed number of students who took aspirin, 177, is equal to the expected number, 177.

   (d) **True / False** In this study, there are two proportion parameters: the proportion who had reduced flu symptoms and the proportion who took aspirin when they had the flu. One of these parameters is *un*known (the proportion who had reduced flu symptoms) and one is known (the proportion who took aspirin when they had the flu).

   (e) The *expected* number of students whose flu symptoms are reduced when an aspirin is taken (assuming flu symptoms are *independent* of whether or not the aspirin is taken) is calculated in the following way:

   $$\text{reduced flu symptoms} \times \text{assumed proportion who took aspirin} = 201 \times 0.5$$

   which is equal to (circle one) **14.1 / 27.5 / 100.5**.

   (f) The expected number of students whose flu symptoms are reduced when an aspirin is *not* taken, is

   $$\text{reduced flu symptoms} \times \text{assumed proportion who did } \textit{not} \text{ take aspirin} = 201 \times 0.5$$

   which is equal to (circle one) **12.5 / 100.5 / 153**.

(g) The expected number of students whose flu symptoms are *not* reduced when an aspirin is taken, is

no reduced flu symptoms $\times$ assumed proportion who took aspirin

which is equal to (circle one) $\mathbf{153 \times 0.5}$ / $\mathbf{\frac{153\times170}{354}}$ / $\mathbf{201 \times 0.5}$.

(h) The expected number of students whose flu symptoms are *not* reduced when a aspirin is *not* taken, is

no reduced flu symptoms $\times$ assumed proportion who did not take aspirin

which is equal to (circle one) $\mathbf{\frac{153\times170}{354}}$ / $\mathbf{\frac{153\times184}{354}}$ / $\mathbf{153 \times 0.5}$.

(i) **True** / **False** The expected table, created from the observed data, is given below.

expected, $\hat{f}$	aspirin	no aspirin	subtotals
flu symptoms reduced	100.5	100.5	201
flu symptoms not reduced	76.5	76.5	153
subtotals	177	177	354

knowing that 50% took aspirin and 50% did not take aspirin.

(j) Consequently, to determine the test statistic,

flu study	observed, f	expected, $\hat{f}$	$\frac{(f-\hat{f})^2}{\hat{f}}$
flu reduced, aspirin given	120	100.5	3.8
flu reduced, aspirin not given	81	100.5	3.8
flu not reduced, aspirin given	50	76.5	5.0
flu not reduced, aspirin not given	103	76.5	5.0

and so $\sum \frac{(f-\hat{f})^2}{\hat{f}} = 3.8 + 3.8 + 5.0 + 5.0 = 17.5$. The observed test statistic, 17.5, is quite large since observed and expected values are quite different and so this tends to indicate that flu symptoms and aspirin are (circle one) **independent** / **dependent** of one another.

2. *Test Statistic Versus Critical Value, Standardized.*

(a) *Statement.* The statement of the test is (circle none, one or more)

i. H_0 : flu symptoms independent of aspirin
versus H_1 : flu symptoms *not* independent of aspirin

ii. H_0 : flu symptoms independent of aspirin
versus H_1 : flu symptoms dependent on aspirin

iii. H_0 : flu symptoms *not* independent of aspirin
versus H_1 : flu symptoms independent of aspirin

(b) *Test.*

flu study	observed, f	expected, $\hat{f}$	$\frac{(f-\hat{f})^2}{\hat{f}}$
flu reduced, aspirin given	120	100.5	3.8
flu reduced, aspirin not given	81	100.5	3.8
flu not reduced, aspirin given	50	76.5	5.0
flu not reduced, aspirin not given	103	76.5	5.0

and so $\sum \frac{(f-\hat{f})^2}{\hat{f}} = 3.8 + 3.8 + 5.0 + 5.0 = 17.5$.
The standardized upper critical value at $\alpha = 0.01$ with

$$\text{number of categories } -1- \text{ number of unknown parameters}$$
$$= C - 1 - q = 4 - 1 - 1 = 2 \text{ df}$$

is (circle one) **3.84** / **6.63** / **9.21**
(Use PRGM INVCHI2 ENTER 2 ENTER 0.99 ENTER)

(c) *Conclusion.* Since the test statistic, 17.5, is larger than the critical value, 9.21, we (circle one) **accept** / **reject** the null hypothesis that whether or not there is a reduction of flu symptoms is *in*dependent on whether or not the aspirin is administered.

3. *P–Value Versus Level of Significance, Standardized.*

(a) *Statement.* The statement of the test is (circle none, one or more)

i. H_0 : flu symptoms independent of aspirin
versus H_1 : flu symptoms *not* independent of aspirin

ii. H_0 : flu symptoms independent of aspirin
versus H_1 : flu symptoms dependent on aspirin

iii. H_0 : flu symptoms *not* independent of aspirin
versus H_1 : flu symptoms independent of aspirin

(b) *Test.* Since the standardized test statistic is $\chi^2 = 17.5$, with 2 df, the p–value is given by

$$\text{p–value} = P(\chi^2 \geq 17.5)$$

which equals (circle one) **0.00** / **0.08** / **0.10**.
(Use 2nd DISTR 2:χ^2cdf(17.5,E99,2).)
The level of significance is 0.01.

(c) *Conclusion.* Since the p–value, 0.00, is smaller than the level of significance, 0.01, we (circle one) **accept** / **reject** the null hypothesis that whether or not there is a reduction of flu symptoms is *in*dependent on whether or not the aspirin is administered.

(6.1) probability distributions

(a) $C = 2$,

exposed?	proportion, π
no (0)	π_1
yes (1)	π_2

(b) $C = 4$,

litter size	proportion, π
1	0.2
2	0.3
3	0.3
4	0.2

(c) $C = 8$,

number leaves	proportion, π
15	0.02
16	0.03
$\vdots$	$\vdots$
22	0.02

(d) $C = 2$,

germinate?	proportion, π
no (0)	0.20
yes (1)	0.80

(e) $C = 9$,

age	blood pressure	proportion, π
young	low	0.01
young	normal	0.16
$\vdots$	$\vdots$	$\vdots$
old	high	0.14

(f) $C = 12$, similar to (g)

(g) $C = 7$, similar to (c)

(6.3) chance of cure for antibiotic

(a) A (two–sided) 90% CI is given by
$\hat{\pi} \pm z(\alpha/2)\hat{\sigma}_{\hat{\pi}} = \hat{\pi} \pm z(\alpha/2)\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} = 0.86 \pm z(0.10/2)\sqrt{\frac{0.86(0.14)}{100}} =$ $(0.8024, 0.9176)$
90% *confident* chance of cure is in (0.8024, 0.9176)

(b) *P–Value Versus Level of Significance, Standardized.*

1. *Statement.* $H_0 : \pi = 0.85$ versus $H_1 : \pi < 0.85$
2. *Test.* The standardized observed proportion is

$$\text{z test statistic } = \frac{\hat{\pi} - \pi}{\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}} = \frac{0.86 - 0.85}{\sqrt{\frac{0.86(0.14)}{100}}} = 0.2857$$

The p–value, the chance the standardized observed proportion is 0.86 or less, guessing the population proportion is 0.85, is given by

$$\text{p–value} = P(Z \leq 0.2857) = 0.60$$

(Use 2nd DISTR 2:normalcdf(-E99,0.2857).)
The level of significance is given by $\alpha = 0.05$ (say).

3. *Conclusion.* Since the p–value, 0.60, is *greater* than the level of significance, $\alpha = 0.05$, we **accept** the null guess of 0.85.

(6.5) sensitivity

(a) estimate of sensitivity: $\hat{\pi} = 13/15 = 0.8667$, chance test will correctly give a positive result (that the cancer is present)

(b) looking in C.8, at $n - f = 15 - 13 = 2$, gives 95% CI (0.5954, 0.9834), or one minus (0.0166, 0.4046)

(c) *P–Value Versus Level of Significance, Standardized.*

1. *Statement.* $H_0 : \pi = 0.80$ versus $H_1 : \pi > 0.80$
2. *Test.* The p–value, the chance the observed proportion is $\frac{13}{15}$ths or more, guessing the population proportion is 0.80, is given by

$$\text{p–value} = Pr\left(\hat{\pi} \geq \frac{13}{15}\right) = 0.398$$

   (Use 2nd DISTR A:binomcdf(15,0.8,12) and then subtract the result from one: 1− 2nd ANS ENTER.)
   The level of significance is given by $\alpha = 0.05$ (say).
3. *Conclusion.* Since the p–value, 0.398, is greater than the level of significance, $\alpha = 0.05$, we **accept** the null guess of 0.80.

(6.9) peas

(a) A (two–sided) 90% CI is given by
$\hat{\pi} \pm z(\alpha/2)\hat{\sigma}_{\hat{\pi}} = \hat{\pi} \pm z(\alpha/2)\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} = 0.55 \pm z(0.05/2)\sqrt{\frac{0.55(0.45)}{100}} = (0.4810, 0.6910)$
95% *confident* chance round yellow pea is in (0.4810, 0.6910)

(b) *Test Statistic Versus Critical Value, Standardized.*

1. *Statement.* $H_0 : \pi_1 = 9/16, \pi_2 = 3/16, \pi_3 = 3/16, \pi_3 = 1/16$
   versus H_1 : proportions different
2. *Test.* Using,

age	f	$\hat{f}$	$\frac{(f-\hat{f})^2}{\hat{f}}$
round and yellow	110	112.5	0.06
round and green	40	37.5	0.17
angular and yellow	42	37.5	0.54
angular and green	8	12.5	1.62

   the observed test statistic is $\chi^2 = \sum \frac{(f-\hat{f})^2}{\hat{f}} = 2.3823$.
   The standardized upper critical value at $\alpha = 0.05$ with

   $$\text{number of age categories} - 1 = 4 - 1 = 3$$

   degrees of freedom, is 7.8
   (Use PRGM INVCHI2 ENTER 3 ENTER 0.95 ENTER)
3. *Conclusion.* Since the test statistic, 2.4, is smaller than the critical value, 7.8, we accept the null hypothesis that the observed pea distribution is the same as the expected pea distribution.

(6.13) leaf and stems

(a) Since

$$\hat{\pi}_1 - \hat{\pi}_2 = \frac{83}{100} - \frac{33}{100} = 0.50$$

and

$$\hat{\sigma}_{\hat{\pi}} = \sqrt{\frac{0.83(1-0.83)}{100} + \frac{0.33(1-0.33)}{100}} = 0.0602$$

a lower 95% CI is given by
$\hat{\theta} - z(\alpha/2)\hat{\sigma}_{\hat{\theta}} = 0.50 - z(0.05)(0.0602) = 0.4010$

(b) 95% *confident* that there is a proportion of at least 0.4010 more leaf and stem infested plants in the untreated population than in the treated one (eg. treatment works)

(c) *P–Value Versus Level of Significance, Standardized.*

1. *Statement.* $H_0 : \pi = 0.80$ versus $H_1 : \pi \neq 0.80$
2. *Test.* The standardized observed proportion is

$$\text{z test statistic } = \frac{\hat{\pi} - \pi}{\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}} = \frac{0.83 - 0.80}{\sqrt{\frac{0.83(0.17)}{100}}} = 0.79$$

The p–value, the chance the standardized observed proportion is 0.83 or more (or -0.83 or less), guessing the population proportion is 0.80, is given by

$$\text{p–value} = 2 \times P(Z \leq 0.79) = 0.43$$

(Use 2nd DISTR 2:normalcdf(-E99,0.79).)
The level of significance is given by $\alpha = 0.05$ (say).

3. *Conclusion.* Since the p–value, 0.43, is *greater* than the level of significance, $\alpha = 0.05$, we **accept** the null guess of 0.80.

(6.17) germination data given by table below

	germinate (1)	not germinate (2)	total
variety 1	6	4	10
variety 2	3	7	10
total	9	11	20

1. *Statement.* $H_0 : \pi_1 - \pi_2 = 0$ versus $H_1 : \pi_1 - \pi_2 > 0$
2. *Test.* The p–value, the chance the observed number of germinations for variety 1 is greater than or equal to six (6), is given by

$$\text{p–value} = Pr(Y \geq 6) = 1 - Pr(Y \leq 5) = 0.185$$

   (Use PRGM HYPCDF ENTER 20 10 9 5 ENTER and subtract from one (1).)
   The level of significance is given by $\alpha = 0.05$ (say).
3. *Conclusion.* Since the p–value, 0.185, is greater than the level of significance, $\alpha = 0.05$, we accept the null guess of $\pi_1 - \pi_2 = 0$.

(6.19) brands of ice cream

brand A ↓ brand B →	like	dislike	subtotals
like	38	92	130
dislike	13	57	70
subtotals	51	149	200

Since $\hat{\theta} = \frac{92-13}{200} = 0.395$
and $\hat{\sigma}_{\hat{\pi}} = \frac{\sqrt{92+13}}{200} = 0.0512$
a 99% CI is given by $\hat{\theta} \pm z(\alpha/2)\hat{\sigma}_{\hat{\theta}} = 0.395 \pm (2.55)(0.0512) = (0.2644, 0.5256)$
(To determine z(0.005), type 2nd DISTR 3:invNorm(0.995) ENTER)
99% *confident* that there is a proportion of between (0.2644, 0.5256) who like brand A better than brand B (eg. brand A better)

1. *Preliminary Analysis: Figuring Out The Test Statistic.*
   (a) **120**
   (b) **170**
   (c) **177**
   (d) **True**
   (e) **100.5**
   (f) **100.5**
   (g) $\mathbf{153 \times 0.5}$
   (h) $\mathbf{153 \times 0.5}$
   (i) **True**
   (j) **dependent**
2. *Test Statistic Versus Critical Value, Standardized.*
   (a) i, ii
   (b) **9.21**
   (c) **reject**
3. *P–Value Versus Level of Significance, Standardized.*
   (a) i, ii
   (b) **0.00**
   (c) **reject**

You are responsible for *seven* tests and/or confidence intervals for this quiz.

- *Test and Confidence Interval of Proportion (Two Categories), Large Sample.* For a large (typically, $n \geq 30$) random sample, the test statistic for $H_o : \pi = \pi_o$ is

$$Z = \frac{\hat{\pi} - \pi_0}{\hat{\sigma}_{\hat{\pi}}}, \quad \hat{\sigma}_{\hat{\pi}} = \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n}}$$

and the (two–sided) confidence interval is given by,

$$\hat{\pi} \pm z(\alpha/2)\hat{\sigma}_{\hat{\pi}}$$

- *Test and Confidence Interval of Proportion (Two Categories), Small Sample.* For a small random sample, the p–values for left, right and two–sided tests for $H_o : \pi = \pi_o$ are given by, respectively,

$$\begin{aligned}
p_- &= f(0|n, \pi_0) + f(1|n, \pi_0) + \cdots + f(r|n, \pi_0) \\
p_+ &= f(n - r|n, \pi_0) + f(n - r + 1|n, \pi_0) + \cdots + f(n|n, \pi_0) \\
p &= p_- + p_+, \quad \text{where} \\
&\quad p_- = f(0|n, \pi_0) + f(1|n, \pi_0) + \cdots + f(k|n, \pi_0) \\
&\quad p_+ = f(n - k|n, \pi_0) + f(n - k + 1|n, \pi_0) + \cdots + f(n|n, \pi_0) \\
&\quad k = \min(r, n - r)
\end{aligned}$$

where f is a binomial and the confidence interval is determined using table C.8 (page 824) from the text.

- *Goodness of Fit Test (Two or More Categories), Large Sample.* For a large random sample, the test statistic for $H_o : \pi_1 = \pi_{10}, \ldots, \pi_k = \pi_{k0}$, is

$$\chi^2 = \sum \frac{(f - \hat{f})^2}{\hat{f}}$$

The expected number of observations, $\hat{f}$, changes according to the number of specified (known) proportion parameters.

- *Test and Confidence Interval of Difference in Proportions (Independent Populations), Large Sample.* For a large (typically, $n \geq 30$) random sample, the test statistic for $H_o : \theta = \pi_1 - \pi_2 = \theta_0$ is

$$z = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_{\hat{\theta}}}, \quad \hat{\sigma}_{\hat{\theta}} = \sqrt{\frac{\hat{\pi}_1(1 - \hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1 - \hat{\pi}_2)}{n_2}}$$

and the (two–sided) confidence interval is given by,

$$\hat{\theta} \pm z(\alpha/2)\hat{\sigma}_{\hat{\theta}}$$

- *Fisher's Test: Test and Confidence Interval of Difference in Proportions (Independent Populations), Small Sample.* The p–values for left, right and two–sided tests for $H_o : \pi_1 - \pi_2$ are given by, respectively,

$$\begin{aligned} p_- &= f(y|0, n_1, n) + f(y|1, n_1, n) + \cdots + f(y|y, n_1, n) \\ p_+ &= f(n_1|n_1 - y, n_1, n) + f(n_1|n_1 - y + 1, n_1, n) + \cdots + f(n_1|n_1, n_1, n) \\ p &= p_- + p_+, \quad \text{where} \\ & \quad p_- = f(y|0, n_1, n) + f(y|1, n_1, n) + \cdots + f(y|y, n_1, n) \\ & \quad p_+ = f(n_1|n_1 - y, n_1, n) + f(n_1|n_1 - y + 1, n_1, n) + \cdots + f(n_1|n_1, n_1, n) \end{aligned}$$

  where f is a hypergeometric.

- *McNemar's Test: Test and Confidence Interval of Difference in Proportions (Dependent Populations), Large Sample.* For a large (typically, $n \geq 30$) random sample, the test statistic for $H_o : \theta = \pi_1 - \pi_2 = \theta_0$ is

$$z = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_{\hat{\theta}}}, \quad \hat{\theta} = \frac{B - C}{N}, \quad \hat{\sigma}_{\hat{\theta}} = \frac{\sqrt{B + C}}{N}$$

  and the (two–sided) confidence interval is given by,

$$\hat{\theta} \pm z(\alpha/2)\hat{\sigma}_{\hat{\theta}}$$

- *Goodness of Fit Test (Two or More Categories) For Two Populations, Large Sample.* For a large random sample, the test statistic for $H_o : \pi_{11} = \pi_{21}, \ldots, \pi_{1k} = \pi_{2k}$, is

$$\chi^2 = \sum \frac{(f - \hat{f})^2}{\hat{f}}$$

  where it is assumed there are no known proportion parameters.