Nonparametric Statistics (STAT 3504): Exam 1, Fall 2017

Instructor: Robert B. Gramacy

Instructions

1. Please direct all questions concerning this exam only to the instructor. Do not collaborate with a student in the classroom during the exam.

2. Please turn off (not on vibrate) all cellular phones before starting this exam.

3. When you finish your exam, print your name at the bottom of this cover sheet and sign the honor pledge.

4. You are allowed one sheet of notes, front and back, that you produced yourself, a calculator and a writing device.

5. For all calculations on probability distributions, consult the table on the sheet provided.

6. For a two-sided test under a Binomial or HyperGeometric, please calculate the p-value as $\varphi \approx 2 \times \min\{\mathbb{P}(T \leq t), \mathbb{P}(T \geq t)\}$.

7. All hypothesis tests default to an $\alpha = 0.05$ level unless otherwise stated; CIs may specify a different α.

8. For tolerance limits, please use the approximations given in class (with reference to $x_{1-\alpha}$ quantiles). The true values are not computable without R.

I agree to take the exam under the above directions, and have completed this exam in compliance with the Virginia Tech Honor Code.

Name: Key

Signature: ________________

Date: ________________

Problem 1 [20 pts]

The grade-point average of 15 college seniors selected at random from the graduating class is as follows:

1.2	3.9	2.4	2.7	2.9
2.8	3.0	3.8	3.5	2.3
2.8	1.5	2.7	3.8	2.0

Is it reasonable to assume the upper quartile (75th percentile) is 3.5?

a. (3 pts) Indicate the test you will perform, and state the null and alternative hypotheses.

A two-tailed quantile test:

H_0: 75th %-tile is 3.5

H_1: 75th %-tile is not 3.5

b. (5 pts) Compute the test statistic. & provide its null distribution.

$t_1 = 11 \quad \leq 3.5$

$t_2 = 12 \quad < 3.5$

$T_1, T_2 \sim Bin(n=15, p=0.75)$

c. (7 pts) Calculate the p-value.

$$\begin{aligned} p &= 2\min\{P(T_1 \leq t_1),\ P(T_2 \geq t_2)\} \\ &= 2\min\{P(T_1 \leq 11),\ 1-P(T_2 \leq 11)\} \\ &= 2\min\{0.538,\ 1-0.538\} = 0.922 \end{aligned}$$

d. (5 pts) What do you conclude?

fail to reject H_0 — not enough evidence to suggest 3.5 is not the 75% quantile.

Problem 2 [20 pts]

The following data corresponds to the number of municipal bonds sold by a broker during a 12 month period. Can we conclude there is an increasing trend?

16 12 17 14 14 18 20 11 18 12 17 19

a. (3 pts) Indicate the test you will perform, and state the null and alternative hypotheses.

upper-tailed Cox & Stuart test for trend.

H_0: no increasing trend $\equiv P(+) \le P(-)$

H_1: increasing trend $\equiv P(+) > P(-)$

b. (5 pts) Compute the test statistic. & provide the null distribution.

$t = \#$ of $+$'s $= 4$

$n = \#$ nonzero $= 6$

$X = 16 \quad 12 \quad 17 \quad 14 \quad 14 \quad 18$

$Y = 20 \quad 11 \quad 18 \quad 12 \quad 17 \quad 19$

$+ \quad - \quad + \quad - \quad + \quad +$

$T \sim Bin(n=6, p=0.5)$

c. (7 pts) Calculate the p-value.

$$
\begin{aligned}
p = P(T \geq 4) &= 1 - P(T \le 3) \\
&= 1 - 0.65625 \\
&= 0.34375
\end{aligned}
$$

d. (5 pts) What do you conclude?

fail to reject H_0, not enough evidence to ~~[illegible]~~ say there is an increasing trend

Problem 3: [25 pts]

In 1988 the department of Highway Safety and Motor Vehicles in Florida compiled information about the use of safety equipment and fatal injuries in car accidents. It was found that when no safety equipment was used 691 accidents ended with fatal injuries, while 162,527 did not. When a seat belt was used 1609 accidents ended in fatal injuries, while 412,368 did not. Is the probability of having a fatal injury the same whether you use a seat belt or not?

a. (5 pts) Indicate the test you will perform, and state the null and alternative hypotheses.

$p_1 \equiv$ prob fatal injury w/o safety equip ; $p_2 \equiv$ with

$H_0 = p_1 = p_2$ a two-tailed proportion test

$H_1 : p_1 \neq p_2$

b. (8 pts) Compute the test statistic.

	F	NF	T
NS	$O_{11} = 691$	$O_{12} = 162527$	$163218 = n_1$
S	$O_{21} = 1609$	$O_{22} = 412368$	$413977 = n_2$
	$C_1 = 2300$	$C_2 = 574895$	$577195 = N$

$$t_1 = \frac{\sqrt{N}(O_{11}O_{22} - O_{21}O_{12})}{\sqrt{n_1 n_2 c_1 c_2}}$$

$$= 1.884061$$

$T_1 \sim N(0,1)$

c. (8 pts) Calculate the p-value.

$$q = 2 \cdot P(T_1 \geq |t_1|) = 2 \cdot P(T_1 \geq 1.884061)$$

$$= 2 \cdot [1 - P(T_1 \leq 1.884061)]$$

$$= 2 \cdot [1 - 0.9702175] = 0.05955$$

d. (5 pts) What do you conclude?

Fail to reject H_0 — not enough evidence to conclude that prob of fatal injury differs if safety equip deployed or not.

Problem 4: [20 pts]

The following are the numbers of prescriptions filled by to pharmacies over a 22 day period.

Day	A	B		Day	A	B		Day	A	B		Day	A	B	
1	29	15	−	7	14	13	−	13	12	10	−	19	24	18	−
2	11	17	+	8	19	11	−	14	17	11	−	20	12	16	+
3	17	12	−	9	10	24	+	15	15	10	−	21	10	14	+
4	15	12	−	10	28	11	−	16	16	22	+	22	25	17	−
5	14	15	+	11	21	15	−	17	10	17	+				
6	22	16	−	12	23	19	−	18	28	13	−				

Does pharmacy A fill more prescriptions than pharmacy B?

a. (3 pts) Indicate the test you will perform, and state the null and alternative hypotheses.

lower-tailed sign test

$H_0: P(+) \geq P(-)$

$H_1: P(+) < P(-)$

b. (5 pts) Compute the test statistic. & provide its null distribution.

$t = \#$ of $+$'s $= 7$ 22 non-ties

$T \sim Bin(n=22, p=0.5)$

c. (7 pts) Calculate the p-value.

$p = P(T \leq t) = P(T \leq 7) = 0.06690025$

d. (5 pts) What do you conclude?

fail to reject H_0 & conclude that pharmacy A does not fill more prescriptions than pharmacy B

Problem 5: [15 pts]

A mail order catalog company surveyed 146 of its customers to find out the shipping time (from order date to the date of delivery) for their recent orders using regular U.S. mail.

a. (8 pts) At least what percentage of its customers can expect delivery between $X^{(5)}$ and $X^{(144)}$ as given by its sample observations with 95% certainty?

$$n = 146, \quad 1-\alpha = 0.95, \quad r = 5, \quad m = 146 + 1 - 144 = 3$$

$$\chi^2_{0.95,\, 2(r+m)=16} = 26.29623$$

$$q \doteq \frac{4n - 2(r+m-1) - \chi^2_{1-\alpha,\, 2(r+m)}}{4n - 2(r+m-1) + \chi^2_{1-\alpha,\, 2(r+m)}}$$

$$= \frac{4 \cdot 146 - 2(5+3-1) - 26.29623}{4 \cdot 146 - 2(5+3-1) + 26.29623} = 0.9118015$$

b. (7 pts) At least what percentage of its customers can expect delivery between $X^{(5)}$ and $X^{(144)}$ as given by its sample observations with 80% certainty?

$$\chi^2_{0.8,\, 2(r+m)=16} = 20.46508$$

$$q \doteq \frac{4 \cdot 146 - 2(5+3-1) - 20.46508}{4 \cdot 146 - 2(5+3-1) + 20.46508}$$

$$= 0.9306815$$