Formula Card for Exam 4 STA3123

Testing Categorical Probabilities: One – Way Table

Steps

1. Express the Claim:
2. Hypotheses: $H_0: p_1 = p_2 = \ldots = p_k$
   $H_A: \textit{At least}$ one proportion differs significantly *
3. Get Data and determine your alpha level:
4. Calculate the Test Stat: $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$ where $E_i = np_i$
5. Get Your Critical Value: $\chi^2_{\alpha,k-1}$
6. Form Your Initial Conclusion:
7. Final Conclusion:

*Note: hypotheses will vary depending upon the problem.

Testing Categorical Probabilities: Two-Way Table

Steps

1. Express the Claim:

2. Hypotheses: H_0: The two classifications are independent
H_A: The two classifications are dependent

3. Calculate **Expected Cell Values** for each cell:

4. Calculate the Test Stat: $\chi^2 = \sum \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}}$ where $\hat{E}_{ij} = \frac{(\text{row total})(\text{column total})}{n}$

5. Get Your Critical Value: χ^2_α which has (r – 1)(c – 1) degrees of freedom.

6. Form Your Initial Conclusion: Reject the Null if $\chi^2 > \chi^2_\alpha$

7. Final Conclusion:

The Sign Test

Right tailed case:

Step 1 Claim: $\eta > \eta_0$ *

Step 2 Hypotheses: $H_0: \eta \le \eta_0$
$H_A: \eta > \eta_0$

Step 3 Test Stat: $S =$ Number of measurements greater than η_0

Step 4 Determine n: n = sample size minus any sample measurements which have a value equal to η_0.

Step 5 P-value: To find the p-value go to a binomial table in the text appendix, use n from step 4 and p = 0.5, and find $P(X \ge S)$.

$$p - value = P(X \ge S)$$

Step 6 Initial Conclusion: Compare your p-value to alpha, if $p < \alpha$ reject H_0

Step 7 Word Conclusion:

The Sign Test (Left – tailed case):

Step 1 Claim: $\eta < \eta_0$*

Step 2 Hypotheses: $H_0 : \eta \ge \eta_0$
$H_A : \eta < \eta_0$

Step 3 Test Stat: $S = \text{Number of measurements less than } \eta_0$

Step 4 Determine n: n = sample size minus any sample measurements which have a value equal to η_0.

Step 5 P-value: To find the p-value go to a binomial table in the text appendix, use n from step 4 and p = 0.5, and find $P(X \ge S)$.

$$p - value = P(X \ge S)$$

Step 6 Initial Conclusion: Compare your p-value to alpha, if $p < \alpha$ reject H_0

Step 7 Word Conclusion:

The Sign Test (Two – tailed case):

Step 1 Claim: $\eta \ne \eta_0$*

Step 2 Hypotheses: $H_0 : \eta = \eta_0$
$H_A : \eta \ne \eta_0$

Step 3 Test Stat:

$S = \text{Larger of } \mathrm{S_S} (\text{\# of measurements less than } \eta_0) \text{ and } \mathrm{S_B} \ (\text{\# of measurements more than } \eta_0)$

Step 4 Determine n: n = sample size minus any sample measurements which have a value equal to η_0.

Step 5 P-value: To find the p-value go to a binomial table in the text appendix, use n from step 4 and p = 0.5, and find $P(X \ge S)$.

$$p - value = 2P(X \ge S)$$

Step 6 Initial Conclusion: Compare your p-value to alpha, if $p < \alpha$ reject H_0

Step 7 Word Conclusion:

*note: The claim can vary depending upon the wording of the problem.

Large sample case:

Most binomial tables do not exceed n = 25, so when we encounter problems that have a sample size greater than 25 we can use a large sample approximation. The test stat will become:

$z = \dfrac{(S_{\min} + 0.5) - \frac{n}{2}}{\frac{\sqrt{n}}{2}}$, Where $S_{\min}$ is the number of times the less frequent sign occurs.

The Wilcoxon Rank Sum Test (two-tailed case)

Step 1 Claim: D_1 (distribution for pop 1) and D_2 (distribution for pop 2) are identical.

Step 2 H_0 : D_1 and D_2 are identical,

H_A : D_1 is shifted either to the left or to the right of D_2.

Step 3 Rank the data as if it is all one set of values (If ties exist give the tied values the average of the ranks they would have gotten if they were in successive order)

Step 4 Calculate $T_1 =$ sum of the ranks for pop 1 and $T_2 =$ sum of the ranks for pop 2

(As a check make sure $T_1 + T_2 = \frac{n(n+1)}{2}$) If $n_1 < n_2$, $T_1 =$ T = Test stat, If $n_1 > n_2$, $T_2 =$ T = Test stat (if both sample sizes are equal use either sum as your test stat).

Step 5 The rejection region is determined by looking up alpha in the Wilcoxon Rank-Sum table (Reject the null if $T \le T_L$ or $T \ge T_U$)

Step 6 Form your initial conclusion

Step 7 Word final conclusion

Wilcoxon Rank Sum Test

One-tailed test	Two-tailed test
H_0: D_1 and D_2 are identical H_a: D_1 is shifted right of D_2 **or** [H_a: D_1 is shifted left of D_2] ***Test statistic:*** T_1, if $n_1 < n_2$ T_2, if $n_1 > n_2$ Either if $n_1 = n_2$ ***Rejection region:*** T_1: $T_1 \ge T_U$ or [$T_1 \le T_L$] T_2: $T_2 \le T_L$ or [$T_2 \ge T_U$]	H_0: D_1 and D_2 are identical H_a: D_1 is shifted right or left of D_2 ***Test statistic:*** T_1, if $n_1 < n_2$ T_2, if $n_1 > n_2$ Either if $n_1 = n_2$ ***Rejection region:*** $T \le T_L$ or $T \ge T_U$

Wilcoxon Rank Sum Test for Large Samples

- One-tailed test

  H_0: D_1 and D_2 are identical

  H_a: D_1 is shifted right of D_2 **or** H_a: D_1 is shifted left of D_2

  Rejection region:

  $| z | > z_a$

- Two-tailed test

  H_0: D_1 and D_2 are identical

  H_a: D_1 is shifted either right or left of D_2

  Rejection region:

  $| z | > z_{a\ 2}$

Test Statistic: $z = \dfrac{T_1 - \dfrac{n_1(n_1+n_2+1)}{2}}{\sqrt{\dfrac{n_1 n_2 (n_1+n_2+1)}{12}}}$

Wilcoxon Sign-Rank Test (two-tailed case)

Step 1: State the claim.

Step 2: List your hypotheses.

$H_0 : \eta_d = 0$

$H_A : \eta_d \ne 0$

Step 3: Get your ranks—a.) For each pair of data, find the difference (d) by subtracting the second value from the first. Discard any pairs for which d = 0. b.) Take the absolute values of d and rank them from low (= 1) to high. If any differences are tied give the tied differences the average of the ranks they would have if they were in successive order.

Step 4: Add up all the ranks of the negative differences (T_-) and do the same for the ranks of the positive differences (T_+). Let $T = \min(T_+, T_-)$

Step 5: Let n = number of pairs that have a nonzero difference, then find your Critical Value T_0 from the Wilcoxon Sign-Rank Test table.

Step 6: Form your initial conclusion

We reject if our test stat is below our critical value (i.e. – Reject when $T \le T_0$)

Step 7: Word your final conclusion

Wilcoxon Sign-Rank Test (one-tailed case)

Step 1: State the claim.

Step 2: List your hypotheses.

$$\begin{matrix} H_0 : \eta_d \le 0 \\ H_A : \eta_d > 0 \end{matrix} \text{ or } \begin{pmatrix} H_0 : \eta_d \ge 0 \\ H_A : \eta_d < 0 \end{pmatrix}$$

Step 3: Get your ranks—a.) For each pair of data, find the difference (d) by subtracting the second value from the first. Discard any pairs for which d = 0. b.) Take the absolute values of d and rank them from low (= 1) to high. If any differences are tied give the tied differences the average of the ranks they would have if they were in successive order.

Step 4: Add up all the ranks of the negative differences (T_-) [Add up the ranks of the positive differences (T_+)].

Step 5: Let n = number of pairs that have a nonzero difference, then find your Critical Value T_0 from the Wilcoxon Sign-Rank Test table.

Step 6: Form your initial conclusion

We reject if our test stat T_- is below our critical value T_0 [or Reject when $T_+ \le T_0$]

Step 7: Word your final conclusion

The **Kruskal-Wallis *H*-test (Nonparametric CRD)**

Step 1 Express the Claim that all of the treatments produce the same median.

Step 2 Hypotheses

$H_0 : \eta_1 = \eta_2 = \ldots = \eta_k$

H_A : At least two medians differ from each other.

Step 3 Temporarily view the entire data set as a whole and rank the data values (Average the ranks in case of a tie as usual). Then for each treatment add up its ranks.

Step 4 Calculate the test statistic $H = \dfrac{12}{N(N+1)}\left(\dfrac{R_1^2}{n_1} + \dfrac{R_2^2}{n_2} + \cdots + \dfrac{R_k^2}{n_k}\right) - 3(N+1)$

Step 5 Get Critical Value: The H statistic can be approximated by a Chi-Squared distribution with k – 1 degrees of freedom, so we will look up alpha on the Chi-Squared table. The test is always a right tailed test.

Step 6 State your initial conclusion—We will reject the null when $H > \chi^2_{\alpha,k-1}$

Step 7 Word final conclusion

Friedman $F_r - Test$ (Nonparametric RBD)

Step 1 Express the claim that all treatments have the same probability distribution.

Step 2 Hypotheses

H_0 : The probability distributions for the treatments are identical.

H_A : At least two of the distributions differ in location.

Step 3 Rank the data values in each block (Average the ranks in case of a tie as usual). Then for each treatment add up its ranks.

Step 4 Calculate the test statistic $F_r = \dfrac{12}{bk(k+1)}\sum R_j^2 - 3b(k+1)$

Step 5 Get Critical Value: The F_r statistic can be approximated by a Chi-Squared distribution with k – 1 degrees of freedom, so we will look up alpha on the Chi-square table. The test is always a right tailed test.

Step 6 State your initial conclusion—We will reject the null when $F_r > \chi^2_{\alpha,k-1}$

Step 7 Word final conclusion.