Name: Key AP Statistics Review Date:

Test of significance

Complete the 4 essential steps (*Hypothesis, Conditions, Mechanics, Conclusion*) for each of the following significance tests.

1. Test for a proportion (1-proportion z-test)

Hypothesis:

$H_0: p = p_0$

$H_a: p \neq p_0$ (2-sided)

$p < p_0$ (1-sided)

$p > p_0$ (1-sided)

Conditions:

- Random/independent sample
- 10% condition: sample size is less than 10% of the population
- Success/Failure condition: $np_0 \geq 10$, $nq_0 \geq 10$

Mechanics: Conditions met → 1-prop z-Test (proportion)

$$z\text{-statistics} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

$$p\text{-value} = p(z > \# \text{ or } < \#)$$

Conclusion:

- p-value is small, we reject the null hypothese. we have sufficient evidence to conclude H_a (in context!)

OR p-value is large, we fail to reject H_0. we have insufficient evidence to conclude H_a (in context!)

2. Test for a difference between two proportions (2-proportion z-test)

make sure you write out what p_1, p_2 stand for!

Hypothesis: $H_0: p_1 = p_2$ OR $p_1 - p_2 = 0$

$H_a: p_1 \neq p_2$ (2-sided)

$p_1 < p_2$ (1-sided)

$p_1 > p_2$ (1-sided)

$p_1 - p_2 \neq 0$ (2-sided)

< 0 (1-sided)

> 0 (1-sided)

conditions: double the ones for one proportion

For each group:

- Random/independent
- 10% condition
- S/F condition

Independent Group Assumption

- The 2 groups are independent of each other.

conditions met → 2-prop z-test (proportion)

Mechanics: (pooled $\hat{p}$)

$$z\text{-statistics} = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\frac{\hat{p}_{pooled}\hat{q}_{pooled}}{n_1} + \frac{\hat{p}_{pooled}\hat{q}_{pooled}}{n_2}}}$$

$$\hat{p}_{pooled} = \frac{\text{Success}_1 + \text{Success}_2}{n_1 + n_2}$$

$$p\text{-value} = p(z > \text{ or } < \text{ or } \neq \#)$$

Conclusion: Same (in context)

3. Test for a mean (1-sample t-test)

Hypotheses: $H_0: \mu = \#$

$H_a: \mu \neq \#$ ($<$, $>$)

conditions:
- Random/independent
- 10% condition
- Nearly Normal condition
  - Normal population stated
  - $n \geq 30 \Rightarrow$ CLT
  - Look at histogram or Normal probability plot

conditions met → 1-sample t-Test

Mechanics: calculate: $\bar{x}, s, n$; $df = n-1$

$$t\text{-statistics} = \frac{\bar{x}-\mu}{s/\sqrt{n}}$$

$$p\text{-value} = P(t \geq t\text{-statistics})$$

Conclusion: In context!

5. Test for a difference between two means (paired – 1-sample t-test)

Hypotheses:

$H_0: \mu_d = \#$

$H_a: \mu_d \neq \#$ ($<$, $>$)

conditions:
- Paired data
- Random
- 10%
- Nearly Normal

Mechanics:

$$t = \frac{\bar{d}-\mu_d}{\frac{s_d}{\sqrt{n}}}$$

p-value

Conclusion: Same 2 sentences in context.

4. Test for a difference between two means (unpaired – 2-sample t-test) make sure you state what μ_1 & μ_2 stand for!

Hypotheses: $H_0: \mu_1 - \mu_2 = 0$

$H_a: \mu_1 - \mu_2 \neq 0$ ($<$, $>$)

conditions: double the 1-sample conditions

For each group:
- random/independent sample
- 10%
- Nearly Normal

Independent Group Assumption:
Two groups are independent of each other.

conditions met → 2-sample t-test

Mechanics:

$$t\text{-statistics}: \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

($\mu_1-\mu_2$: usually "0")

p-value: $P(t <, > \text{test statistic})$

use calculator to find df!

conclusion: Same 2 sentences in context!

Chi-square test for goodness of fit

(always 1-sided)

Hypothesis: words!

H_0: The (data) is distributed as the expected value states

H_a: The (data) is not distributed as the expected value states.

conditions:
- Counted data condition
- Random condition
- Expected cell frequency condition: each cell at least 5.

conditions met → χ^2-model → chi-square goodness of fit test

Mechanics:

$$\chi^2\text{-statistics} = \sum \frac{(obs - exp)^2}{expected}$$

$df = n - 1$
n = # of categories

state p-value!

Conclusion: same 2 sentences in context.

7. Chi-square test for homogeneity of proportions

(2 or more samples, 1 variable)

Hypotheses:

H_0: The proportion of the (distribution) is the same for each group

H_a: The proportion of the (distribution) is not the same for each group

conditions:
- counted data
- Random / independent
- Expected cell frequency at least 5

$$\text{Expected value for cell} = \frac{(\text{row total})(\text{column total})}{\text{total sample size}}$$

conditions met → chi-square test for homogeneity

Mechanics: $df = (rows - 1)(columns - 1)$

state χ^2-statistic, p-value. (may use calculator)

conclusion: in context!

8. Chi-square test for Independence

(one sample, 2 variables)

Hypothesis: H_0: There is no association between _____ & _____.

H_a: There is association between _____ & _____.

The rest is the same as chi-square test for homogeneity.

9. Test for the slope of least squares regression line

Hypothesis: $H_0: \beta_1 = 0$ (no relationship btw 2 variables)

$H_a: \beta \neq 0$ (there is a relationship btw 2 variables)

Conditions:
- straigh Enough (look at scatterplot)
- independence Condition (check for random samples)
- Does the plot Thicken? (look at residual plot)
- Nearly Normal (Look at histogram)

conditions met → use a regression model for the distribution
→ linear regression t-test.

Mechanics: use calculator!

$$t\text{-statistics} = \frac{b_1 - \beta_1}{SE(b_1)}$$

State p-value

Conclusion: same 2 sentences in context.