Final Exam Review

Introduction to Biostatistics, Spring 2006

Overview

- Topics covered since midterm:
  - Confidence Intervals
  - Proportions
  - Contingency tables
  - ANOVA
  - Correlation/Regression
  - Clinical Trials

Concepts

- Point and Interval Estimation
  - Mean, Proportion, Ratio
    - Based on one sample or difference in one of above estimates between 2 or more groups?
  - What distribution?
    - Z, t, chi-square, F,…
- Making predictions
  - Association
    - Correlation
  - Association and Magnitude
    - Regression Models

Concepts

- Hypothesis Testing…
  - Used Z, t, F, chi-square tests, Fisher's, Mantel-Haenszel Method
    - One vs. two samples vs. Multi-samples
    - Paired vs. independent
    - Degrees of freedom
  - AND used confidence intervals
    - Careful whether looking for 0 or 1 in interval!
      - i.e. Difference vs. Ratio and OR vs. log of OR

General Hints

- What kind of data?
  - Continuous vs. Categorical/Dichotomous
  - One vs. Two Samples vs. Multi-samples
  - Paired vs. Independent
  - Sample size – use exact or approximation methods?
- Read question carefully. Are you asked to:
  1. Calculate an estimate,
  2. Calculate a confidence interval,
  3. Conduct a test, or
  4. Make a prediction?
- Think conceptually
  1. Draw pictures
  2. Consider what makes logical sense in context of problem
- Once eliminated impossible answers, then
  1. Pick out important variables given in problem
  2. Choose appropriate formulas to plug these into

Parametric Test Summary

Test	σ	# of Samples	H_o	Distribution	Text Table
z-test	Known	1	$\mu=\mu_0$	N(0,1)	A.3
1-sample t-test	Unknown	1	$\mu=\mu_0$	$T_{(n-1)}$	A.4
Independent-sample t-test	Unknown	2	$\mu_1=\mu_2$	$T_{(n1+n2-2)}$	A.4
Paired t-test (note: $n_1=n_2$)	Unknown	2*	$\mu_1=\mu_2$	$T_{(n-1)}$ for N=# pairs	A.4

*Equivalent to performing a 1-sample t-test on difference between the pairs.

- Make your own charts – help choose appropriate test
  - Direct yourself to equations in the text
- Practice Exam….

Question 1

1. Consider the following two clinical trials concerning an investigational drug given for the treatment of cancer.

Clinical trial #1: 100 people are administered the drug and 40 of these people respond.
Clinical trial #2: 60 people are administered the drug and 24 of these people respond.

Both experiments have a 40% response rate. If a 95% confidence interval for the response rate is constructed for each clinical trial, then which trial yields a tighter (i.e., narrower) confidence interval?

A. Clinical trial #1
B. Clinical trial #2
C. The confidence intervals are the same length
D. The answer cannot be determined without additional information

Power & Sample Size

- To increase power // increase confidence (i.e. tighter interval):
  1. Increase α
     - Q1: Same for both
  2. Increase sample size
     - Q1: Larger in CT #1
  3. Decrease σ
     - Q1: Same for both (since proportions both = 0.4)

Question 1

1. Consider the following two clinical trials concerning an investigational drug given for the treatment of cancer.

Clinical trial #1: 100 people are administered the drug and 40 of these people respond.
Clinical trial #2: 60 people are administered the drug and 24 of these people respond.

Both experiments have a 40% response rate. If a 95% confidence interval for the response rate is constructed for each clinical trial, then which trial yields a tighter (i.e., narrower) confidence interval?

A. Clinical trial #1
B. Clinical trial #2
C. The confidence intervals are the same length
D. The answer cannot be determined without additional information

Question 2

2. The percentage of ideal body weight was determined for 18 randomly selected insulin-dependent diabetics.

A summary is presented below:

Variable | Obs Mean Std. Err. [90% Conf. Interval]

Weight | 18 112.7778 3.399887 ? ?

What is the **90%** confidence interval of the mean percentage of ideal body weight among insulin-dependent diabetics represented by this sample?

A. (106.86, 118.69)
B. (105.60, 119.95)
C. (89.49, 141.06)
D. (89.04, 136.52)

90% CI for Dist'n of Means

$$\bar{x} \pm t_{n-1,1-\alpha} s / \sqrt{n} =$$

$$112.78 \pm 1.740(3.4) =$$

$$(106.86, 118.69)$$

Question 2

2. The percentage of ideal body weight was determined for 18 randomly selected insulin-dependent diabetics.

A summary is presented below:

Variable | Obs Mean Std. Err. [90% Conf. Interval]

Weight | 18 112.7778 3.399887 ? ?

What is the **90%** confidence interval of the mean percentage of ideal body weight among insulin-dependent diabetics represented by this sample?

A. (106.86, 118.69)
B. (105.60, 119.95)
C. (89.49, 141.06)
D. (89.04, 136.52)

Question 3

3. A study is conducted on a group of HIV-infected individuals that exhibit signs of the newly-defined Ascending Neuromuscular Weakness Syndrome (ANWS). A p-value is calculated (using a 1-sample t-test) to test the null hypothesis that the mean lactate level was equal to five (versus an alternative hypothesis that the mean lactate level was not equal to 5). This yielded a p-value of 0.074. If one was to construct a 2-sided 90% confidence interval and a 2-sided 95% confidence interval for the mean lactate level in this population, then:

A. Both confidence intervals would include 5.
B. Neither confidence interval would include 5.
C. The 95% confidence interval would include 5; however the 90% confidence interval would not.
D. The 90% confidence interval would include 5; however the 95% confidence interval would not.

- Again, 90% CI narrower than 95% CI (reflects α).
- $0.05 < 0.074 < 0.10$

Question 3

3. A study is conducted on a group of HIV-infected individuals that exhibit signs of the newly-defined Ascending Neuromuscular Weakness Syndrome (ANWS). A p-value is calculated (using a 1-sample t-test) to test the null hypothesis that the mean lactate level was equal to five (versus an alternative hypothesis that the mean lactate level was not equal to 5). This yielded a p-value of 0.074. If one was to construct a 2-sided 90% confidence interval and a 2-sided 95% confidence interval for the mean lactate level in this population, then:

A. Both confidence intervals would include 5.
B. Neither confidence interval would include 5.
C. The 95% confidence interval would include 5; however the 90% confidence interval would not.
D. The 90% confidence interval would include 5; however the 95% confidence interval would not.

The table below summarizes information from a clinical trial that compared two treatments (a test drug and placebo) for a respiratory disorder. Subjects were then observed for a favorable or unfavorable response.

Treatment	Favorable	Unfavorable	Total
Placebo	16	48	64
Test Drug	40	20	60
Total	56	68	124

Pearson chi-square test (1 d.f.) = 21.709 p-value = 0.001

Question 4

4. The null and alternative hypotheses associated with the Pearson chi-square test is:

A. H0: Treatment is associated with response, HA: no association
B. H0: Treatment is associated with better response, HA: no association
C. H0: No association between treatment and response, HA: treatment is associated with better response
D. H0: No association between treatment and response, HA: treatment is associated with response

Treatment	Favorable	Unfavorable	Total
Placebo	16	48	64
Test Drug	40	20	60
Total	56	68	124

Pearson chi-square test (1 d.f.) = 21.709 p-value = 0.001

- If no difference (i.e. treatment not associated with response), we would **expect** proportions in favorable vs. unfavorable cells to be equal between placebo and test drug groups.
- Pearson chi-Square test compares differences between observed and expected.

Question 4

4. The null and alternative hypotheses associated with the Pearson chi-square test is:

A. H0: Treatment is associated with response, HA: no association
B. H0: Treatment is associated with better response, HA: no association
C. H0: No association between treatment and response, HA: treatment is associated with better response
D. H0: No association between treatment and response, HA: treatment is associated with response

Question 5

5. The conclusion obtained from the above chi-square test is:

A. Reject the null hypothesis and conclude no association
B. Reject the null hypothesis and conclude that there is evidence of association
C. Do not reject the null hypothesis and conclude that there is no evidence to suggest association
D. Reject the null hypothesis and conclude a linear relationship

Treatment	Favorable	Unfavorable	Total
Placebo	16	48	64
Test Drug	40	20	60
Total	56	68	124

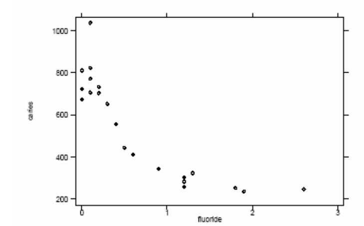
Pearson chi-square test (1 d.f.) = 21.709 p-value = 0.001

Question 5

5. The conclusion obtained from the above chi-square test is:

A. Reject the null hypothesis and conclude no association
B. Reject the null hypothesis and conclude that there is evidence of association
C. Do not reject the null hypothesis and conclude that there is no evidence to suggest association
D. Reject the null hypothesis and conclude a linear relationship

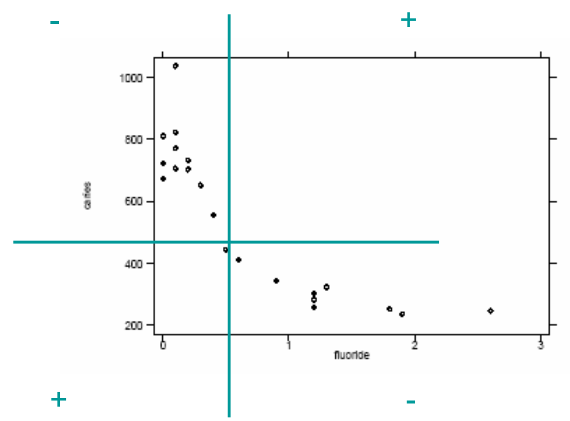
Consider the following scatter plot depicting the relationship between the number of dental caries (number of caries per 100 children) and the level of fluoridation of the public water supply (fluoride parts per million).



Question 6

6. What is the general relationship of fluoride level and the number of caries is (r is the correlation coefficient)?

A. Strongly negative (r<0)
B. Strongly positive (r>0)
C. Strongly negative (r>0)
D. There is no relationship (r.0)



Question 6

6. What is the general relationship of fluoride level and the number of caries is (r is the correlation coefficient)?

A. Strongly negative (r<0)
B. Strongly positive (r>0)
C. Strongly negative (r>0)
D. There is no relationship (r.0)

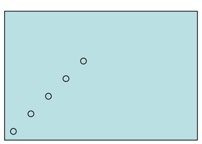
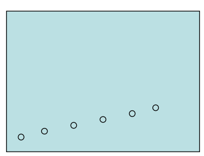
Question 7

7. Consider the relationship between two continuous variables and the two-dimensional scatterplot associated with these variables. If the correlation coefficient is –1, then

A. All points lie in a straight line with a slope of –1.
B. All points lie in a straight line with an unknown negative slope.
C. All points do not lie in a straight line but the best fitting regression line has a slope of –1.
D. There is a strong positive relationship between these two variables.

Correlation

- Always between -1 and 1.
- Only exactly -1 or 1 if on a straight line.
- Tells us Association, NOT magnitude (ie slope)!

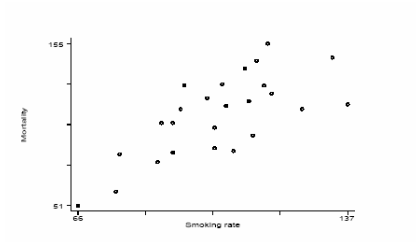
- Same correlation!
- Regression extends correlation to assess magnitude
  - As well as multiple variables simultaneously

Question 7

7. Consider the relationship between two continuous variables and the two-dimensional scatterplot associated with these variables. If the correlation coefficient is –1, then

A. All points lie in a straight line with a slope of –1.
B. All points lie in a straight line with an unknown negative slope.
C. All points do not lie in a straight line but the best fitting regression line has a slope of –1.
D. There is a strong positive relationship between these two variables.

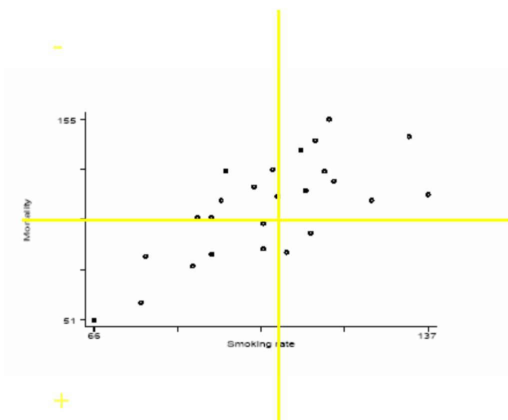
Consider the following graph of mortality rates versus smoking rates from an English occupational health study. The units are expressed as relative rates to the average of the entire population (100%). So a rate of 120% would be 20% higher than the population average, while a rate of 80% would be 20% lower than average.



Question 8

8. Based on the scatterplot, it appears that:

A. Smoking rate is uncorrelated with mortality rate
B. Smoking rate is positively correlated with mortality rate
C. Smoking rate is negatively correlated with mortality rate
D. It can be concluded that smoking causes higher mortality



Question 8

8. Based on the scatterplot, it appears that:

A. Smoking rate is uncorrelated with mortality rate
B. Smoking rate is positively correlated with mortality rate
C. Smoking rate is negatively correlated with mortality rate
D. It can be concluded that smoking causes higher mortality

The STATA output of the simple linear regression of mortality versus smoking rates is presented below.

```
  Source |       SS       df       MS              Number of obs =      25
---------+------------------------------           F(  1,    23) =   24.23
   Model |  8395.74904     1  8395.74904           Prob > F      =  0.0001
Residual |  7970.25096    23   346.53265           R-squared     =  0.5130
---------+------------------------------           Adj R-squared =  0.4918
   Total |    16366.00    24  681.916667           Root MSE      =  18.615

------------------------------------------------------------------------------
    mort |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
 smoking |   1.087532    .2209452      4.922   0.000       .6304724    1.544592
   _cons |  -2.885319    23.03372     -0.125   0.901       -50.5342    44.76356
------------------------------------------------------------------------------
```

Question 9

9. Does an increase in the smoking rate significantly increase the mortality rate?

A. Yes, since the coefficient is 1.08 (>0) and the p-value is 0.000.

B. No, since the estimate of the constant is –2.88 and is not significant.

C. Yes, since R-squared = 0.5130.

D. No, an increase in the smoking rate decreases the mortality rate.

Regression

- Response = Mortality (y_i)
- Predictor = Smoking Rate (x_i)
- Coefficient = Slope (β_1)
- Intercept = β_0

$$y_i = \beta_0 + \beta_1 x_i + e$$


```
  Source |       SS       df       MS              Number of obs =      25
---------+------------------------------           F(  1,    23) =   24.23
   Model |  8395.74904     1  8395.74904           Prob > F      =  0.0001
Residual |  7970.25096    23   346.53265           R-squared     =  0.5130
---------+------------------------------           Adj R-squared =  0.4918
   Total |    16366.00    24  681.916667           Root MSE      =  18.615

------------------------------------------------------------------------------
    mort |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
 smoking |   1.087532    .2209452      4.922   0.000       .6304724    1.544592
   _cons |  -2.885319    23.03372     -0.125   0.901       -50.5342    44.76356
------------------------------------------------------------------------------
```


Question 9

9. Does an increase in the smoking rate significantly increase the mortality rate?

A. Yes, since the coefficient is 1.08 (>0) and the p-value is 0.000.

B. No, since the estimate of the constant is –2.88 and is not significant.

C. Yes, since R-squared = 0.5130.

D. No, an increase in the smoking rate decreases the mortality rate.

Question 10

10. As a result of the regression, a 10% relative increase in smoking comparatively will result in:

A. a 10.9% relative *increase* in mortality
B. a 28.8% relative *decrease* in mortality
C. a 28.8% relative *increase* in mortality
D. a 10.9% relative *decrease* in mortality

10 x 1.087 = 10.87

Question 10

10. As a result of the regression, a 10% relative increase in smoking comparatively will result in:

A. a 10.9% relative *increase* in mortality
B. a 28.8% relative *decrease* in mortality
C. a 28.8% relative *increase* in mortality
D. a 10.9% relative *decrease* in mortality

An experiment was conducted on rodents (*rattus norvegicus*) to determine if the drug haldol influenced the amount of "rough-play". Five groups were included in the experiment. Subsequent behavior in terms of mean "rough-play" behaviors was recorded. The graph below displays the mean rough-play differences from baseline (i.e., after the drug was administered compared to activity *before* drug administration).

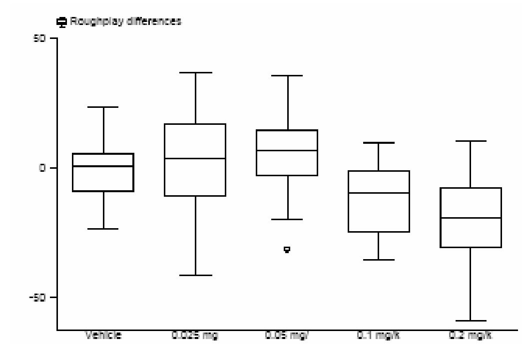
Summary of Roughplay differences

Haldol dose	Mean	Std. Dev.	Freq.
Vehicle	-.63158419	13.290618	19
0.025 mg	1.1481501	20.447041	18
0.05 mg/	5.0499951	16.035879	20
0.1 mg/k	-12.754389	13.783537	19
0.2 mg/k	-20.55	16.431064	20
Total	-5.6631968	18.558994	96

Analysis of Variance

Source	SS	df	MS	F Test-Statistic	Prob > F (p-value)
Between groups	8999.32832	4	2249.83208	8.63	0.0000
Within groups	23722.1178	91	260.682613		
Total	32721.4461	95	344.436275		

Bartlett's test for equal variances: chi2(4) = 4.1710 Prob>chi2 = 0.383



Question 11

11. The null and alternative hypothesis of this test is:

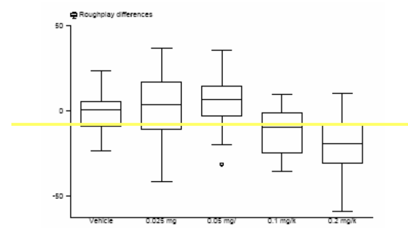
A. H0: β=0 (there is no linear association between dose and activity levels) versus H1: β≠0 (there is a linear association between dose and activity levels)

B. H0: :1=:2=:3=:4=:5 versus H1: At least one dose group is different from at least one of the others

C. H0: There is no association between dose and activity versus H1: There is an association (as measured by the chi-square test)

D. H0: 0 = β versus H1: 0 ≠ β , i.e., there is no linear association between dose and activity level versus there is a linear association

ANOVA



- Testing to see if there's any difference between at least two of these groups.
- Look at the overlap between individual 'boxes'.
  - If less variation within groups, as compared to between groups, then we should see less overlap and more differentiation.

Question 11

11. The null and alternative hypothesis of this test is:

A. H0: 0 = β (there is no linear association between dose and activity levels) versus H1: 0 ≠ β (there is a linear association between dose and activity levels)

B. H0: :1=:2=:3=:4=:5 versus H1: At least one dose group is different from at least one of the others

C. H0: There is no association between dose and activity versus H1: There is an association (as measured by the chi-square test)

D. H0: 0 = β versus H1: 0 ≠ ρ , i.e., there is no linear association between dose and activity level versus there is a linear association

Question 12

12. The appropriate test, p value and the conclusion is as follows:

A. An *F* test, with 1 and 94 degrees of freedom, p=0.000<0.05, reject the null hypothesis

B. A chi-square test with 4 degrees of freedom, p=0.383>0.05, do not reject the null hypothesis

C. A t test with 91 degrees of freedom, p=0.000<0.05, reject the null hypothesis

D. An *F* test with 4, and 91 degrees of freedom, p=0.000<0.05, reject the null hypothesis

```
             |   Summary of Roughplay differences
 Haldol dose |        Mean   Std. Dev.       Freq.
-------------+------------------------------------
     Vehicle |  -.63158419   13.290618          19
    0.025 mg |   1.1481501   20.447041          18
    0.05 mg/ |   5.0499951   16.035879          20
    0.1 mg/k |  -12.754389   13.783537          19
    0.2 mg/k |      -20.55   16.431064          20
-------------+------------------------------------
       Total |  -5.6631968   18.558994          96

                        Analysis of Variance

    Source              SS         df      MS            F Test-    Prob > F
                                                         Statistic  (p-value)
------------------------------------------------------------------------------
Between groups      8999.32832      4   2249.83208        8.63      0.0000
 Within groups      23722.1178     91   260.682613
------------------------------------------------------------------------------
    Total           32721.4461     95   344.436275

Bartlett's test for equal variances:  chi2(4) =   4.1710  Prob>chi2 = 0.383
```

Question 12

12. The appropriate test, p value and the conclusion is as follows:

A. An *F* test, with 1 and 94 degrees of freedom, p=0.000<0.05, reject the null hypothesis

B. A chi-square test with 4 degrees of freedom, p=0.383>0.05, do not reject the null hypothesis

C. A t test with 91 degrees of freedom, p=0.000<0.05, reject the null hypothesis

D. An *F* test with 4, and 91 degrees of freedom, p=0.000<0.05, reject the null hypothesis

ANOVA, cont.

- Next step would be to find WHERE those differences are…
- Post-hoc tests – pairwise comparisons
  - Bonferonni
  - Tukey
  - Scheffe

Consider the tables of individuals stratified by sex, investigating the association of smoking status and aortic stenosis:

Males

Aortic Stenosis	*Smoking Status* *Yes*	*No*	*Total*
Yes	37	25	62
No	24	20	44
Total	61	45	106

Females

Aortic Stenosis	*Smoking Status* *Yes*	*No*	*Total*
Yes	14	29	43
No	19	47	66
Total	33	76	109

```
             sex |       OR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               0 |   1.233333     .5681868   2.678358     5.660377 (Cornfield)
               1 |   1.194192     .5252581   2.719303     5.055046 (Cornfield)
-----------------+-------------------------------------------------
           Crude |   1.471576     .8578159   2.524475              (Cornfield)
    M-H combined |   1.214868     .6877312    2.14605
-----------------+-------------------------------------------------
Test for heterogeneity (M-H)      chi2(1) =     0.003  Pr>chi2 = 0.9558

                  Test that combined OR = 1:
                          Mantel-Haenszel chi2(1) =      0.45
                                          Pr>chi2 =    0.5044
```

Question 13

13. Considering the confidence intervals of the individual odds ratios, test the hypotheses (separately) that there is *no increased* risk among male and female smokers to develop aortic stenosis versus non-smokers.

A. In both cases we can conclude that there is an increased risk (reject Ho)

B. We conclude that there is no increased risk evident in either group (cannot reject Ho)

C. There is an increased risk among male smokers but not among female smokers

D. Do not smoke!

```
             sex |       OR      [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               0 |   1.233333     .5681868   2.678358     5.660377 (Cornfield)
               1 |   1.194192     .5252581   2.719303     5.055046 (Cornfield)
-----------------+-------------------------------------------------
           Crude |   1.471576     .8578159   2.524475              (Cornfield)
     M-H combined |   1.214868     .6877312    2.14605
-----------------+-------------------------------------------------
Test for heterogeneity (M-H)      chi2(1) =      0.003  Pr>chi2 = 0.9558

                   Test that combined OR = 1:
                                Mantel-Haenszel chi2(1) =      0.45
                                                Pr>chi2 =    0.5044
```

- No difference if OR=1.0…
- 1.0 falls in both confidence intervals!

Question 13

13. Considering the confidence intervals of the individual odds ratios, test the hypotheses (separately) that there is *no increased* risk among male and female smokers to develop aortic stenosis versus non-smokers.

A. In both cases we can conclude that there is an increased risk (reject Ho)

B. We conclude that there is no increased risk evident in either group (cannot reject Ho)

C. There is an increased risk among male smokers but not among female smokers

D. Do not smoke!

Question 14

14. Given the results of the "test for heterogeneity" the appropriate course of action is to

A. Proceed with the Mantel-Haenszel (M-H) test

B. Disregard the M-H test as inappropriate

C. Get more data

D. Perform separate analyses

```
             sex |       OR       [95% Conf. Interval]   M-H Weight
-----------------+-------------------------------------------------
               0 |   1.233333      .5681868   2.678358     5.660377 (Cornfield)
               1 |   1.194192      .5252581   2.719303     5.055046 (Cornfield)
-----------------+-------------------------------------------------
           Crude |   1.471576      .8578159   2.524475              (Cornfield)
    M-H combined |   1.214868      .6877312    2.14605
-----------------+-------------------------------------------------
Test for heterogeneity (M-H)     chi2(1) =     0.003  Pr>chi2 = 0.9558

                  Test that combined OR = 1:
                              Mantel-Haenszel chi2(1) =      0.45
                                              Pr>chi2 =    0.5044
```

- M-H chi-square P-value of 0.9558.
- Fail to reject the test for heterogeneity ➔ the odds ratios are the same across strata (sex).

Question 14

14. Given the results of the "test for heterogeneity" the appropriate course of action is to

A. Proceed with the Mantel-Haenszel (M-H) test

B. Disregard the M-H test as inappropriate

C. Get more data

D. Perform separate analyses

- We could then use the M-H combined odds ratio
  - Don't need to stratify by sex – no effect modification

A Christmas tree survey was conducted to determine if there was a difference between urban and rural tree users in terms of artificial tree use. The results are as follows:

Group	Total Tree Users	Number with Artificial Trees
Urban	261	89
Rural	160	64

Question 15

15. The 90% Confidence interval for the difference between the proportions (urban minus rural) of artificial tree users is:

A. (-0.211, 0.129)

B. (0.023, 0.441)

C. (-0.139, 0.021)

D. (-0.246, 0.234)

90% C.I. for Difference in Proportions

$$\hat{p}_1 = \frac{89}{261} = 0.34 \quad \hat{p}_2 = \frac{64}{160} = 0.40$$

$$\hat{p}_1 - \hat{p}_2 \pm 1.645\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$= -0.06 \pm 1.645\sqrt{\frac{0.34(0.66)}{261} + \frac{0.4(0.6)}{160}}$$

$$= (-0.139, 0.021)$$

Question 15

15. The 90% Confidence interval for the difference between the proportions (urban minus rural) of artificial tree users is:

A. (-0.211, 0.129)
B. (0.023, 0.441)
C. (-0.139, 0.021)
D. (-0.246, 0.234)

Question 16

16. The resulting interpretation of the above confidence interval is:

A. We are confident that the two groups differ with respect to tree preference.
B. We did not find enough evidence to suggest that the two groups differ with respect to tree preference.
C. We found enough evidence to conclude that there is no difference between the two groups with respect to tree preference.
D. There is an association between population and tree preference.

- Zero falls in 90% confidence interval, so will fall in 95% confidence interval too!

Question 16

16. The resulting interpretation of the above confidence interval is:

A. We are confident that the two groups differ with respect to tree preference.
B. We did not find enough evidence to suggest that the two groups differ with respect to tree preference.
C. We found enough evidence to conclude that there is no difference between the two groups with respect to tree preference.
D. There is an association between population and tree preference.

An educator wishes to compare a new directed reading activity in the classroom versus a standard curriculum in helping elementary school children improve some aspects of their reading ability. A third-grade class of 21 students follows these activities for 8 weeks while a control group of 23 children follow a standard curriculum. Students were then given a test designed to measure aspects of reading ability. The results are as follows:

Group	N	Mean	Standard deviation
Treatment	21	51.48	11.01
Control	23	41.52	17.15

Question 17

17. The 95% confidence interval for the mean difference (treatment minus control) is

A. (9.96, 27.94)
B. (0.97, 18.95)
C. (0, 18.95)
D. (-18.95, 18.95)

95% C.I. for Difference in Means

- Independent groups, unequal variances
- Estimated difference:
  51.48 - 41.52 = 9.96
- Don't need to do full CI calculations: The interval should be symmetric around 9.96!
- Still good to know the formula on page 272!

Question 17

17. The 95% confidence interval for the mean difference (treatment minus control) is

A. (9.96, 27.94)
B. (0.97, 18.95)
C. (0, 18.95)
D. (-18.95, 18.95)

Question 18

18. Thus the resulting p-value of a test of the null hypothesis of equal improvement between the two groups versus the alternative hypothesis that the mean improvement in the two groups is different is:

A. >0.05
B. >0.10
C. < 0.05
D. One cannot determine anything concerning the p-value without more information

- Zero doesn't fall in 95% confidence interval!

Question 18

18. Thus the resulting p-value of a test of the null hypothesis of equal improvement between the two groups versus the alternative hypothesis that the mean improvement in the two groups is different is:

A. >0.05
B. >0.10
C. < 0.05
D. One cannot determine anything concerning the p-value without more information

Question 19

19. In a study of the effect of smoking on lung cancer, it is estimated that the odds ratio (for a subject that smokes relative to a subject that does not smoke) is 2.6. This means that:

A. A person with lung cancer is 2.6 times as likely to be a smoker as a non-smoker.
B. A person that smokes is 2.6 times as likely to develop lung cancer as a person that does not smoke.
C. A person that smokes has a 2.6% higher chance of developing lung cancer than a person that does not smoke.
D. Lung cancer increases the prevalence of smoking by 2.6%.

Odds Ratios

- Measure of relative risk
  - Also use RD, RR to compare risks
- Typically used in case-control studies
- Ranges from 0 to infinity (skewed)
  - Why we use log transformation to create symmetry for CI's… (Don't forget to convert back!)
- Odds=P/(1-P)
- Odds Ratio=

$$\frac{P(\text{Cancer}\mid\text{Smoker})/(1-P(\text{Cancer}\mid\text{Smoker}))}{P(\text{Cancer}\mid\text{Non})/(1-P(\text{Cancer}\mid\text{Non}))}$$

$$= 2.6$$

Question 19

19. In a study of the effect of smoking on lung cancer, it is estimated that the odds ratio (for a subject that smokes relative to a subject that does not smoke) is 2.6. This means that:

A. A person with lung cancer is 2.6 times as likely to be a smoker as a non-smoker.

B. A person that smokes is 2.6 times as likely to develop lung cancer as a person that does not smoke.

C. A person that smokes has a 2.6% higher chance of developing lung cancer than a person that does not smoke.

D. Lung cancer increases the prevalence of smoking by 2.6%.

Question 20

20. You are reading a journal article and you see a p-value of 0.11. Which of the following are true?

A. Since the p-value is not significant, there is no need to investigate further.

B. You should determine to what hypotheses the p-value corresponds so that the result can be put into context.

C. P-values are sensitive to sample size and thus you should determine upon what sample size the p-value is based.

D. B and C

- Always assess P-values in context!

Question 20

20. You are reading a journal article and you see a p-value of 0.11. Which of the following are true?

A. Since the p-value is not significant, there is no need to investigate further.

B. You should determine to what hypotheses the p-value corresponds so that the result can be put into context.

C. P-values are sensitive to sample size and thus you should determine upon what sample size the p-value is based.

D. B and C

Question 21

21. Biostatistics may be used to:

A. Describe the results of medical, public health, environmental, and ecological studies

B. Isolate and estimate the effect of new treatments for treating diseases while controlling for other important variables and identify risk factors for various diseases and syndromes

C. Make decisions about public health policy and plan for new treatments and interventions

D. All of the above

Question 21

21. Biostatistics may be used to:

A. Describe the results of medical, public health, environmental, and ecological studies

B. Isolate and estimate the effect of new treatments for treating diseases while controlling for other important variables and identify risk factors for various diseases and syndromes

C. Make decisions about public health policy and plan for new treatments and interventions

D. All of the above

GOOD LUCK!!!!