

# BIOS 600: Principles of Statistical Inference

## Case Studies in Hypothesis Testing

Fall 2012

# Reading

- ▶ Flowchart on selecting a test procedure (Sakai)

## Case Study: PKU

Fifteen patients with confirmed phenylketonuria (PKU) were identified and placed on dietary therapy before 4 months of age. The children were given an IQ test upon starting grade school, and their unaffected siblings closest in age were also given the test. The differences in IQ (PKU-unaffected) of the siblings are: 12, -12, 22, -24, 6, -13, -25, 2, 22, -1, -23, -22, -4, 6, and 8.

Researchers wish to know whether PKU children have the same IQ on average as their unaffected siblings.

## Case Study: PKU

How do you determine which is the most appropriate test to use?  
The data are graphed below.



The following slides show results of several different hypothesis tests. Pick the single most appropriate test for these data, specify the null and alternative hypotheses, and interpret the results.

# PKU Data: Sign Test

```
. bitesti 15 6 .5
```

N	Observed k	Expected k	Assumed p	Observed p
<b>15</b>	<b>6</b>	<b>7.5</b>	<b>0.50000</b>	<b>0.40000</b>
<hr/>				
Pr(k >= 6)		= 0.849121	(one-sided test)	
Pr(k <= 6)		= 0.303619	(one-sided test)	
Pr(k <= 6 or k >= 9)		= 0.607239	(two-sided test)	

# PKU Data: Wilcoxon Signed-Rank Test

```
. signrank PKU=Sibling
```

Wilcoxon signed-rank test

sign	obs	sum ranks	expected
positive	6	40.5	60
negative	9	79.5	60
zero	0	0	0
all	15	120	120

unadjusted variance      310.00

adjustment for ties      -0.75

adjustment for zeros      0.00

---

adjusted variance      309.25

Ho: PKU = Sibling

z = -1.109

Prob > |z| = 0.2675

# PKU Data: t Test

```
. ttest PKU=Sibling
```

Paired t test

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
PKU	15	94.66667	4.789539	18.5498	84.39413	104.9392
Sibling	15	98.8	3.450604	13.36413	91.39919	106.2008
diff	15	-4.133333	4.10559	15.90088	-12.93895	4.672282

mean(diff) = mean(PKU - Sibling) t = -1.0068  
Ho: mean(diff) = 0 degrees of freedom = 14

Ha: mean(diff) < 0 Ha: mean(diff) != 0 Ha: mean(diff) > 0  
Pr(T < t) = 0.1656 Pr(|T| > |t|) = 0.3311 Pr(T > t) = 0.8344

## Case Study: Apnea

A new drug is being pilot tested for treatment of apnea (a transient cessation of breathing) in premature infants. The outcome of interest is the average number of apneic episodes per hour and was measured before and after drug treatment (an average was taken over several hours for each infant).

Investigators would like to know whether the new drug has any effect on the average number of hourly episodes. Which test is most appropriate, and why? Specify the null and alternative hypotheses for the test you choose.

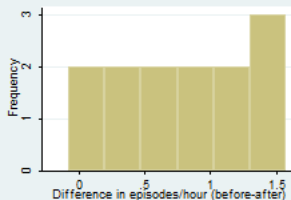
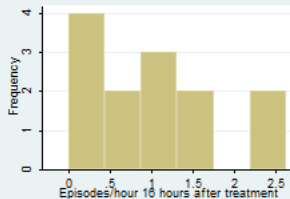
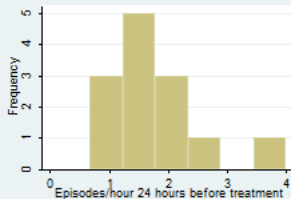


# Apnea Data

. list before after diff

	<b>before</b>	<b>after</b>	<b>diff</b>
1.	1.71	.13	1.58
2.	1.25	.88	.37
3.	2.13	1.38	.7500001
4.	1.29	.13	1.16
5.	1.58	.25	1.33
6.	4	2.63	1.37
7.	1.4	1.38	.02
8.	1.08	.5	.58
9.	1.83	1.25	.58
10.	.67	.75	-.08
11.	1.1	0	1.1
12.	2.71	2.38	.3299999
13.	1.96	1.13	.83

## Apnea Data



# DDT

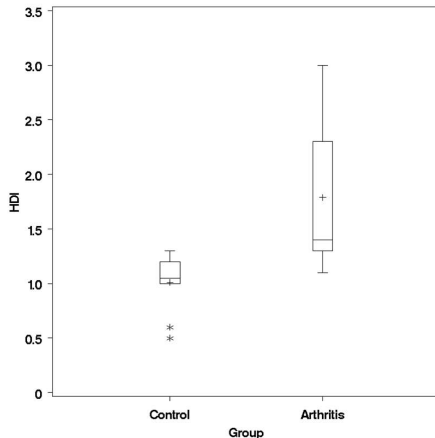
Suppose you are interested in knowing whether mean blood levels of DDE, a metabolite of the pesticide DDT, are the same for women with breast cancer and for women who are healthy controls. For each woman with breast cancer, one control was recruited to the study, matched to the cancer patient on age, menopausal status, and date.

171 women with breast cancer were recruited, and the differences between women with cancer and without (cancer DDE minus control DDE) had sample mean  $\bar{d} = 2.7\text{ng/ml}$  and standard deviation  $s = 15.9\text{ ng/ml}$ . Which test would you use to determine whether this is a statistically significant difference and why? Provide the null and alternative hypotheses for your test.

If you were to conduct a t-test, what would the test statistic be?

## Case Study: Rheumatoid Arthritis

Below are boxplots of the heat distribution index (HDI), a measure of disease activity in rheumatoid arthritis.



# Rheumatoid Arthritis

The data come from two independent groups of subjects: a group with diagnosed rheumatoid arthritis, and a healthy control group. What method would you use to analyze the data

- ▶ if you were told there were 10 subjects in each group?
- ▶ if you were told there were 200 subjects in each group?
- ▶ Suppose instead you have three groups: control, rheumatoid arthritis, and osteoarthritis. Now what method would you use to analyze the data, and why?

## Anxiety in Students

A researcher wishes to compare the average anxiety levels during final exams of four student populations: high school students, college undergraduates, graduate students, and medical students. The researcher randomly selects 50 students from each of these populations and administers a questionnaire designed to measure state anxiety. She fits an ANOVA model and obtains  $F = 6.1$  with corresponding  $p < 0.001$ .

- ▶ What are the null and alternative hypotheses of the test?
- ▶ What are the numerator and denominator degrees of freedom?
- ▶ What do you conclude based on the test result?
- ▶ What would be your next steps?

# Important Concepts

- ▶ z-scores and the standard normal distribution

# Important Concepts

- ▶ z-scores and the standard normal distribution
- ▶ Central Limit Theorem and sampling distribution of the mean



# Important Concepts

- ▶ z-scores and the standard normal distribution
- ▶ Central Limit Theorem and sampling distribution of the mean
- ▶ **Parameters versus statistics**

# Important Concepts

- ▶ z-scores and the standard normal distribution
- ▶ Central Limit Theorem and sampling distribution of the mean
- ▶ Parameters versus statistics
- ▶ Standard deviation versus standard error

# Important Concepts

- ▶ z-scores and the standard normal distribution
- ▶ Central Limit Theorem and sampling distribution of the mean
- ▶ Parameters versus statistics
- ▶ Standard deviation versus standard error
- ▶ Point and interval estimation

# Important Concepts

- ▶ z-scores and the standard normal distribution
- ▶ Central Limit Theorem and sampling distribution of the mean
- ▶ Parameters versus statistics
- ▶ Standard deviation versus standard error
- ▶ Point and interval estimation
- ▶ What is a confidence interval?

# Important Concepts

- ▶ z-scores and the standard normal distribution
- ▶ Central Limit Theorem and sampling distribution of the mean
- ▶ Parameters versus statistics
- ▶ Standard deviation versus standard error
- ▶ Point and interval estimation
- ▶ What is a confidence interval?
- ▶ What factors make a confidence interval wider? More narrow?

## Important Concepts

- ▶ How does the t-distribution compare to the normal distribution?

## Important Concepts

- ▶ How does the t-distribution compare to the normal distribution?
- ▶ When and why is the t-distribution used instead of the normal distribution?

# Important Concepts

- ▶ How does the t-distribution compare to the normal distribution?
- ▶ When and why is the t-distribution used instead of the normal distribution?
- ▶ What are degrees of freedom?



# Important Concepts

- ▶ How does the t-distribution compare to the normal distribution?
- ▶ When and why is the t-distribution used instead of the normal distribution?
- ▶ What are degrees of freedom?
- ▶ Hypothesis testing framework

# Important Concepts

- ▶ How does the t-distribution compare to the normal distribution?
- ▶ When and why is the t-distribution used instead of the normal distribution?
- ▶ What are degrees of freedom?
- ▶ Hypothesis testing framework
- ▶ Null and alternative hypotheses

# Important Concepts

- ▶ How does the t-distribution compare to the normal distribution?
- ▶ When and why is the t-distribution used instead of the normal distribution?
- ▶ What are degrees of freedom?
- ▶ Hypothesis testing framework
- ▶ Null and alternative hypotheses
- ▶ Difference between one-sided and two-sided tests

# Important Concepts

- ▶ How does the t-distribution compare to the normal distribution?
- ▶ When and why is the t-distribution used instead of the normal distribution?
- ▶ What are degrees of freedom?
- ▶ Hypothesis testing framework
- ▶ Null and alternative hypotheses
- ▶ Difference between one-sided and two-sided tests
- ▶  $\alpha$

# Important Concepts

- ▶ How does the t-distribution compare to the normal distribution?
- ▶ When and why is the t-distribution used instead of the normal distribution?
- ▶ What are degrees of freedom?
- ▶ Hypothesis testing framework
- ▶ Null and alternative hypotheses
- ▶ Difference between one-sided and two-sided tests
- ▶  $\alpha$
- ▶ Power

# Important Concepts

- ▶ How does the t-distribution compare to the normal distribution?
- ▶ When and why is the t-distribution used instead of the normal distribution?
- ▶ What are degrees of freedom?
- ▶ Hypothesis testing framework
- ▶ Null and alternative hypotheses
- ▶ Difference between one-sided and two-sided tests
- ▶  $\alpha$
- ▶ Power
- ▶ Type I and type II errors

# Important Concepts

- ▶ When to reject  $H_0$

# Important Concepts

- ▶ When to reject  $H_0$
- ▶ What a p-value really means



# Important Concepts

- ▶ When to reject  $H_0$
- ▶ What a p-value really means
- ▶ How the t statistic is constructed

# Important Concepts

- ▶ When to reject  $H_0$
- ▶ What a p-value really means
- ▶ How the t statistic is constructed
- ▶ How to evaluate  $H_0$  using a confidence interval

## Important Concepts

- ▶ When to reject  $H_0$
- ▶ What a p-value really means
- ▶ How the t statistic is constructed
- ▶ How to evaluate  $H_0$  using a confidence interval
- ▶ How is power affected if we change the sample size, minimum detectable difference,  $\alpha$ , sidedness of test, standard deviation, etc.?

# Important Concepts

- ▶ When to reject  $H_0$
- ▶ What a p-value really means
- ▶ How the t statistic is constructed
- ▶ How to evaluate  $H_0$  using a confidence interval
- ▶ How is power affected if we change the sample size, minimum detectable difference,  $\alpha$ , sidedness of test, standard deviation, etc.?
- ▶ How is sample size affected if we change the power, minimum detectable difference,  $\alpha$ , sidedness of test, standard deviation, etc.?

# Important Concepts

- ▶ What are the different types of t-tests, and when do we use each?

# Important Concepts

- ▶ What are the different types of t-tests, and when do we use each?
- ▶ When do we use ANOVA, and when is it not valid?

# Important Concepts

- ▶ What are the different types of t-tests, and when do we use each?
- ▶ When do we use ANOVA, and when is it not valid?
- ▶ What are common nonparametric alternatives to t-tests and ANOVA, and when do we use them?

# Important Concepts

- ▶ What are the different types of t-tests, and when do we use each?
- ▶ When do we use ANOVA, and when is it not valid?
- ▶ What are common nonparametric alternatives to t-tests and ANOVA, and when do we use them?
- ▶ What are multiple comparisons, and why are they a problem?