Exam 3 Lecture Review

Statistics 121
2019

Topics covered

- Exploratory data analysis for C to Q and C to C relationships (slides 87-88, 103-105)
- Sampling distribution of p-hat (slides 4-6)
- One sample z- procedures for proportions (slides 7-30)
- Matched pairs t procedures for means (slides 112-149)
- Two sample t procedures for means (slides 103-109)
- ANOVA (slides 31-61)
- Two sample z procedures for proportions (slides 99-102, 152-153)
- Two way tables, conditional distributions (slides 89-98)
- Chi-square test (slides 62-86)

Symbols

μ: Mean of a population; also Mean of the sampling distribution of $\bar{x}$ ("mean of $\bar{x}$")

$\bar{x}$: Mean of a sample

p: Proportion of a population; also Mean of the sampling distribution of $\hat{p}$ ("mean of $\hat{p}$")

$\hat{p}$: Proportion of a sample

σ: Standard deviation of a population

s: Standard deviation of a sample

n: Sample size

$\frac{\sigma}{\sqrt{n}}$: Standard deviation of the sampling distribution of $\bar{x}$ (also called "std. dev. of $\bar{x}$")

$\frac{s}{\sqrt{n}}$: Standard error of $\bar{x}$; estimates standard deviation of the sampling distribution of $\bar{x}$

$t^*\frac{s}{\sqrt{n}}$: Margin of error for estimating μ

$\sqrt{\frac{p(1-p)}{n}}$: Standard deviation of the sampling distribution of $\hat{p}$ (also called std. dev. of $\hat{p}$)

$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$: Standard error of $\hat{p}$; estimates standard deviation of the samp. dist. of $\hat{p}$

$z * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$: Margin of error for estimating p

1. Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used? Assume the proportion of infected trees in the orchard is 10%.

- Describe the sampling distribution of p-hat in this context.

- Mean of the sampling distribution:
- Standard deviation of the sampling distribution:
- Shape of the sampling distribution:

1. Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

Why was the sampling distribution of $\hat{p}$ approximately Normal assuming p=0.10?

a. Because n = 120 > 30.
b. Because 120(0.10) = 12 > 10 and 120(0.90) = 108 > 10
c. Since 10 trees are infected, $\hat{p}$ =0.0833 and 120(0.0833) = 10 ≥ 20 and 120(0.91667) = 110 > 10.

1.

Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

Why was the sampling distribution of $\hat{p}$ approximately Normal assuming p=0.10?

a. Because n = 120 > 40.
b. Because 120(0.10) = 12 > 10 and 120(0.90) = 108 > 10
c. Since 10 trees are infected, $\hat{p}$ =0.0833 and 120(0.0833) = 10 ≥ 20 and 120(0.91667) = 110 > 10.

Note:

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Ex3. Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

What symbol should be used in the hypotheses?

a. μ

b. p

c. $\overline{x}$

d. $\hat{p}$

Ex3. Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

What symbol should be used in the hypotheses?

a. μ

b. p

c. $\bar{x}$

d. $\hat{p}$

Ex3 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

What hypotheses should be tested?

a. H_0: $p = 0.10$ versus H_a: $p < 0.10$

b. H_0: $\hat{p} = 0.10$ versus H_a: $\hat{p} < 0.10$

c. H_0: $p = 0.10$ versus H_a: $p > 0.10$

d. H_0: $\hat{p} = 0.10$ versus H_a: $\hat{p} > 0.10$

Ex3 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

What hypotheses should be tested?

a. H_0: $p = 0.10$ versus H_a: $p < 0.10$

b. H_0: $\hat{p} = 0.10$ versus H_a: $\hat{p} < 0.10$

c. H_0: $p = 0.10$ versus H_a: $p > 0.10$

d. H_0: $\hat{p} = 0.10$ versus H_a: $\hat{p} > 0.10$

NOTE: Answers “b” and “d” are always wrong because we never use statistic symbols in our hypotheses.

Ex3 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

Which hypothesis corresponds to using the cheaper spray?

a. H_0: $p = 0.10$
b. H_a: $p < 0.10$

Ex3 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

Which hypothesis corresponds to using the cheaper spray?

a. H_0: $p = 0.10$
b. H_a: $p < 0.10$

NOTE: The standard spray will be used unless we reject H_0.

Ex3 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

H_0: $p = 0.10$ –use standard spray
H_a: $p < 0.10$ –use cheaper spray

What is the Type I error for these hypotheses?

a. Deciding to use the cheaper spray when should.
b. Deciding to use the cheaper spray when shouldn't
c. Sticking with standard spray when cheaper was ok.
d. Sticking with standard spray when should.

Ex3 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

H_0: $p = 0.10$ –use standard spray
H_a: $p < 0.10$ –use cheaper spray

What is the Type I error for these hypotheses?

a. Deciding to use the cheaper spray when should.
b. Deciding to use the cheaper spray when shouldn't
c. Sticking with standard spray when cheaper was ok.
d. Sticking with standard spray when should.

Ex3 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

H_0: $p = 0.10$ –use standard spray
H_a: $p < 0.10$ –use cheaper spray

What is the Type II error for these hypotheses?

a. Deciding to use the cheaper spray when should.
b. Deciding to use the cheaper spray when shouldn't
c. Sticking with standard spray when cheaper was ok.
d. Sticking with standard spray when should.

Ex3 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

H_0: $p = 0.10$ –use standard spray
H_a: $p < 0.10$ –use cheaper spray

What is the Type II error for these hypotheses?

a. Deciding to use the cheaper spray when should.
b. Deciding to use the cheaper spray when shouldn't
c. Sticking with standard spray when cheaper was ok.
d. Sticking with standard spray when should.

Ex3 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

H_0: $p = 0.10$ –use standard spray
H_a: $p < 0.10$ –use cheaper spray

What is power for these hypotheses?

a. Probability of using the cheaper spray when should.
b. Probability of using the cheaper spray when shouldn't
c. Probability of sticking with standard spray when cheaper was ok.
d. Probability of sticking with standard spray when should.

Ex3 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

H_0: $p = 0.10$ –use standard spray
H_a: $p < 0.10$ –use cheaper spray

What is power for these hypotheses?

a. Probability of using the cheaper spray when should.
b. Probability of using the cheaper spray when shouldn't
c. Probability of sticking with standard spray when cheaper was ok.
d. Probability of sticking with standard spray when should.

Ex3 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

Describe the sampling distribution of $\hat{p}$ assuming H_0 is true.

a. Somewhat skewed to the right with mean = $\hat{p}$ and standard deviation unknown.

b. Approximately Normal with mean = $\hat{p}$ and standard deviation = $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

c. Approximately Normal with mean = 0.10 and

standard deviation = $\sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{(0.10)(0.90)}{120}} = 0.0274$

EX2 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

Describe the sampling distribution of $\hat{p}$ assuming H_0 is true.

a. Somewhat skewed to the right with mean = $\hat{p}$ and standard deviation unknown.

b. Approximately Normal with mean = $\hat{p}$ and standard deviation = $\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

c. Approximately Normal with mean = 0.10 and standard deviation = $\sqrt{\dfrac{p_0(1-p_0)}{n}} = \sqrt{\dfrac{(0.10)(0.90)}{120}} = 0.0274$

EX2 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

What is the standard error of $\hat{p}$?

a. $\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

b. $\sqrt{\dfrac{p_0(1-p_0)}{n}}$

EX2 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

What is the standard error of $\hat{p}$?

a. $\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

b. $\sqrt{\dfrac{p_0(1-p_0)}{n}}$

Ex3 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

Are the necessary conditions met so that the standard Normal curve can be used to obtain the *P*-value?

a. Yes, because trees will be randomly selected and sampling distribution of $\hat{p}$ is approximately Normal.
b. No, because trees are not randomly allocated to treatments.
c. No, because the sample size is not large enough for Normality.

Ex3 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

Are the necessary conditions met so that the standard Normal curve can be used to obtain the *P*-value?

a. Yes, because trees will be randomly selected and sampling distribution of $\hat{p}$ is approximately Normal.
b. No, because trees are not randomly allocated to treatments.
c. No, because the sample size is not large enough for Normality.

Ex3 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

Suppose 10 trees in the sample are infected. What is the *P*-value?

Ex3 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

Suppose 10 trees in the sample are infected. What is the *P*-value?

$$\hat{p} = \tfrac{10}{120} = 0.0833$$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.0833 - 0.10}{\sqrt{\frac{0.10(1-0.10)}{120}}} = \frac{-0.01667}{0.0274} = -0.608$$

We look up z = -0.61 and get *P*-value = 0.2709

H_0: $p = 0.10$ versus H_a: $p < 0.10$

Ex3 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. Should the cheaper spray be used?

Interpret the *P*-value = 0.27 in context.

a. There is a 27% probability that the null hypothesis is true.
b. If the proportion of infected trees were 0.10, the probability of obtaining a sample proportion of 0.0833 or smaller is 0.2709.
c. The chances of rejecting a false null hypothesis is 1 – 0.27 = 0.73 or 73%
d. Assuming that the percentage of infected trees is 10%, the probability of obtaining a sample result of 8.3% or more infected trees is 27%.

$H_0: p = 0.10$ versus $H_a: p < 0.10$

Ex3 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. Should the cheaper spray be used?

Interpret the *P*-value = 0.27 in context.

a. There is a 27% probability that the null hypothesis is true.
b. If the proportion of infected trees were 0.10, the probability of obtaining a sample proportion of 0.0833 or smaller is 0.2709.
c. The chances of rejecting a false null hypothesis is 1 – 0.27 = 0.73 or 73%
d. Assuming that the percentage of infected trees is 10%, the probability of obtaining a sample result of 8.3% or more infected trees is 27%.

H_0: $p = 0.10$ versus H_a: $p < 0.10$

Ex3 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

Using *P*-value = 0.2709 and $\alpha = 0.05$, should the cheaper spray be used?

a. Yes, because we fail to reject the null hypothesis.
b. Yes, because we reject the null hypothesis.
c. No, because we fail to reject the null hypothesis.
d. No, because reject the null hypothesis.

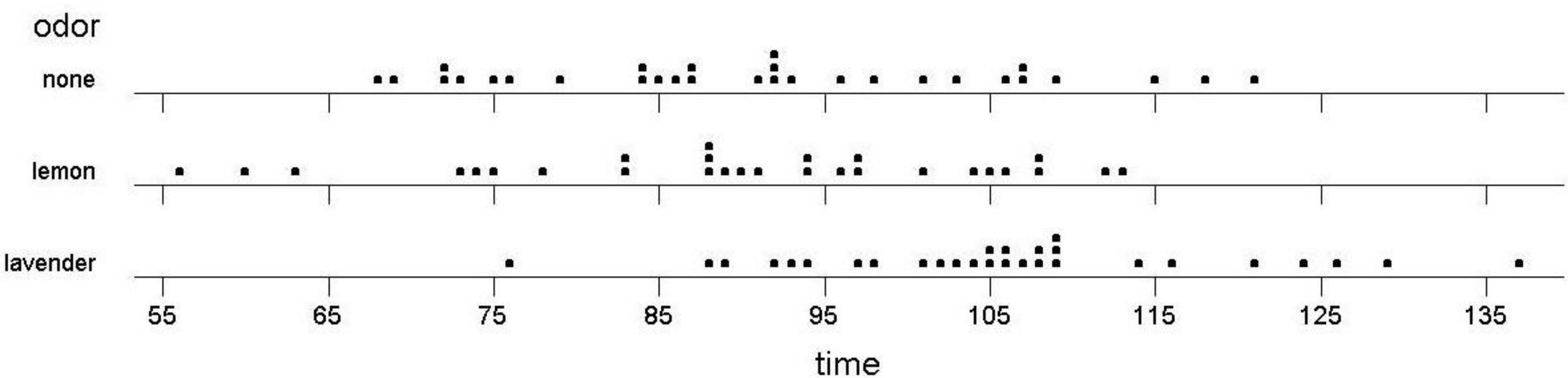
H_0: $p = 0.10$ versus H_a: $p < 0.10$

Ex3 Suppose a standard spray will be used in an orchard unless the supervisor can show that the proportion of infected trees is less than 10%, in which case a cheaper, less effective spray will be used. 120 trees will be randomly sampled. Should the cheaper spray be used?

Using *P*-value = 0.2709 and $\alpha = 0.05$, should the cheaper spray be used?

a. Yes, because we fail to reject the null hypothesis.
b. Yes, because we reject the null hypothesis.
c. No, because we fail to reject the null hypothesis.
d. No, because reject the null hypothesis.

2. **Do good smells bring good business?** A study was conducted at a small pizza restaurant in France on Saturday evenings in May. On one evening a relaxing lavender odor was spread through the restaurant; on another, a stimulating lemon odor; and a third evening had no odor. Treatments were randomly assigned to days. On each evening the time (in minutes) customers stayed in the restaurant was recorded.



2 **Do good smells bring good business?** A study was conducted at a small pizza restaurant in France on Saturday evenings in May. On one evening a relaxing lavender odor was spread through the restaurant; on another, a stimulating lemon odor; and a third evening had no odor. Treatments were randomly assigned to days. On each evening the time (in minutes) customers stayed in the restaurant was recorded.

Is this an experiment or an observational study? Can we conclude causation?

a. experiment—can conclude causation
b. experiment—cannot conclude causation
c. observational study—can conclude causation
d. observational study—cannot conclude causation

2 **Do good smells bring good business?** A study was conducted at a small pizza restaurant in France on Saturday evenings in May. On one evening a relaxing lavender odor was spread through the restaurant; on another, a stimulating lemon odor; and a third evening had no odor. Treatments were randomly assigned to days. On each evening the time (in minutes) customers stayed in the restaurant was recorded.

Is this an experiment or an observational study? Can we conclude causation?

a. experiment—can conclude causation
b. experiment—cannot conclude causation
c. observational study—can conclude causation
d. observational study—cannot conclude causation

2 **Do good smells bring good business?** A study was conducted at a small pizza restaurant in France on Saturday evenings in May. On one evening a relaxing lavender odor was spread through the restaurant; on another, a stimulating lemon odor; and a third evening had no odor. Treatments were randomly assigned to days. On each evening the time (in minutes) customers stayed in the restaurant was recorded.

What is the response variable? What type of variable?

a. Time spent in restaurant—categorical
b. Time spent in restaurant—quantitative
c. Type of odor—categorical
d. Type of odor—quantitative

2 **Do good smells bring good business?** A study was conducted at a small pizza restaurant in France on Saturday evenings in May. On one evening a relaxing lavender odor was spread through the restaurant; on another, a stimulating lemon odor; and a third evening had no odor. Treatments were randomly assigned to days. On each evening the time (in minutes) customers stayed in the restaurant was recorded.

What is the response variable? What type of variable?

a. Time spent in restaurant—categorical
b. Time spent in restaurant—quantitative
c. Type of odor—categorical
d. Type of odor—quantitative

2 **Do good smells bring good business?** A study was conducted at a small pizza restaurant in France on Saturday evenings in May. On one evening a relaxing lavender odor was spread through the restaurant; on another, a stimulating lemon odor; and a third evening had no odor. Treatments were randomly assigned to days. On each evening the time (in minutes) customers stayed in the restaurant was recorded.

What is the explanatory variable? What type of variable?

a. Time spent in restaurant—categorical
b. Time spent in restaurant—quantitative
c. Type of odor—categorical
d. Type of odor—quantitative

2 **Do good smells bring good business?** A study was conducted at a small pizza restaurant in France on Saturday evenings in May. On one evening a relaxing lavender odor was spread through the restaurant; on another, a stimulating lemon odor; and a third evening had no odor. Treatments were randomly assigned to days. On each evening the time (in minutes) customers stayed in the restaurant was recorded.

What is the explanatory variable? What type of variable?

a. Time spent in restaurant—categorical
b. Time spent in restaurant—quantitative
c. Type of odor—categorical
d. Type of odor—quantitative

2 <u>Do good smells bring good business?</u> A study was conducted at a small pizza restaurant in France on Saturday evenings in May. On one evening a relaxing lavender odor was spread through the restaurant; on another, a stimulating lemon odor; and a third evening had no odor. Treatments were randomly assigned to days. On each evening the time (in minutes) customers stayed in the restaurant was recorded.

What is the population of interest?

a. all customers
b. all times spent in restaurant
c. all evenings
d. all restaurants

2 **Do good smells bring good business?** A study was conducted at a small pizza restaurant in France on Saturday evenings in May. On one evening a relaxing lavender odor was spread through the restaurant; on another, a stimulating lemon odor; and a third evening had no odor. Treatments were randomly assigned to days. On each evening the time (in minutes) customers stayed in the restaurant was recorded.

What is the population of interest?

a. all customers
b. all times spent in restaurant
c. all evenings
d. all restaurants

2 **Do good smells bring good business?** A study was conducted at a small pizza restaurant in France on Saturday evenings in May. On one evening a relaxing lavender odor was spread through the restaurant; on another, a stimulating lemon odor; and a third evening had no odor. Treatments were randomly assigned to days. On each evening the time (in minutes) customers stayed in the restaurant was recorded.

What is the population of interest?

a. all customers
b. all times spent in restaurant
c. all evenings
d. all restaurants

2 **Do good smells bring good business?** A study was conducted at a small pizza restaurant in France on Saturday evenings in May. On one evening a relaxing lavender odor was spread through the restaurant; on another, a stimulating lemon odor; and a third evening had no odor. Treatments were randomly assigned to days. On each evening the time (in minutes) customers stayed in the restaurant was recorded.

What is the parameter(s) of interest?

a. times spent in restaurant
b. all restaurant customers
c. mean number of minutes spent in restaurant for each odor treatment
d. proportion of customers who stayed longer than average in restaurant

2 **Do good smells bring good business?** A study was conducted at a small pizza restaurant in France on Saturday evenings in May. On one evening a relaxing lavender odor was spread through the restaurant; on another, a stimulating lemon odor; and a third evening had no odor. Treatments were randomly assigned to days. On each evening the time (in minutes) customers stayed in the restaurant was recorded.

What is the parameter(s) of interest?

a. times spent in restaurant
b. all restaurant customers
c. mean number of minutes spent in restaurant for each odor treatment
d. proportion of customers who stayed longer than average in restaurant

2 **Do good smells bring good business?** A study was conducted at a small pizza restaurant in France on Saturday evenings in May. On one evening a relaxing lavender odor was spread through the restaurant; on another, a stimulating lemon odor; and a third evening had no odor. Treatments were randomly assigned to days. On each evening the time (in minutes) customers stayed in the restaurant was recorded.

What statistical procedure should be performed to answer the research question?

a. two-sample *t*- test for means
b. chi-square test
c. ANOVA
d. matched pairs test

2 **Do good smells bring good business?** A study was conducted at a small pizza restaurant in France on Saturday evenings in May. On one evening a relaxing lavender odor was spread through the restaurant; on another, a stimulating lemon odor; and a third evening had no odor. Treatments were randomly assigned to days. On each evening the time (in minutes) customers stayed in the restaurant was recorded.

What statistical procedure should be performed to answer the research question?

a. two-sample *t*- test for means
b. chi-square test
c. ANOVA
d. matched pairs test

Analysis of Variance results:

Source	df	SS	MS	F-Stat	P-value
Odor type	2	4687	2343	11.08	0.000
Error	84	17760	211		
Total	86	22447			

Factor means:

```
                                             Individual 95% CIs For Mean
                                             Based on Pooled StDev
Level         n       Mean       StDev     -----+---------+---------+---------+-
lavender     29     106.07       13.18                          (------*-----)
lemon        28      89.79       15.44     (------*------)
none         30      91.27       14.93       (------*------)
                                           -----+---------+---------+---------+-
Pooled StDev =      14.54                     88.0      96.0      104.0     112.0
```

2 **Do good smells bring good business?** A study was conducted at a small pizza restaurant in France on Saturday evenings in May. On one evening a relaxing lavender odor was spread through the restaurant; on another, a stimulating lemon odor; and a third evening had no odor. Treatments were randomly assigned to days. On each evening the time (in minutes) customers stayed in the restaurant was recorded.

State the null hypothesis for testing equality of means.

a. $H_0: \bar{x}_1 = \bar{x}_2 = \bar{x}_3$

b. $H_0: \mu_1 = \mu_2, \mu_1 = \mu_3, \mu_2 = \mu_3$

c. $H_0: \mu_1 = \mu_2 = \mu_3$

2 **Do good smells bring good business?** A study was conducted at a small pizza restaurant in France on Saturday evenings in May. On one evening a relaxing lavender odor was spread through the restaurant; on another, a stimulating lemon odor; and a third evening had no odor. Treatments were randomly assigned to days. On each evening the time (in minutes) customers stayed in the restaurant was recorded.

State the null hypothesis for testing equality of means.

a. $H_0: \bar{x}_1 = \bar{x}_2 = \bar{x}_3$

b. $H_0: \mu_1 = \mu_2, \mu_1 = \mu_3, \mu_2 = \mu_3$

c. $H_0: \mu_1 = \mu_2 = \mu_3$

While all the null hypotheses in b and c are mathematically equal, we only state the null hypothesis as given in “c”.

2 **Do good smells bring good business?** A study was conducted at a small pizza restaurant in France on Saturday evenings in May. On one evening a relaxing lavender odor was spread through the restaurant; on another, a stimulating lemon odor; and a third evening had no odor. Treatments were randomly assigned to days. On each evening the time (in minutes) customers stayed in the restaurant was recorded.

State the alternative hypothesis for testing equality of means.

a. $H_0: \bar{x}_1 \neq \bar{x}_2 \neq \bar{x}_3$
b. $H_a: \mu_1 \neq \mu_2 \neq \mu_3$
c. H_a: At least one μ_i differs from the others.

2 **Do good smells bring good business?** A study was conducted at a small pizza restaurant in France on Saturday evenings in May. On one evening a relaxing lavender odor was spread through the restaurant; on another, a stimulating lemon odor; and a third evening had no odor. Treatments were randomly assigned to days. On each evening the time (in minutes) customers stayed in the restaurant was recorded.

State the alternative hypothesis for testing equality of means.

a. $H_0: \bar{x}_1 \neq \bar{x}_2 \neq \bar{x}_3$

b. $H_a: \mu_1 \neq \mu_2 \neq \mu_3$

c. H_a: At least one μ_i differs from the others.

2 [**Do good smells bring good business?**]() Do customers stay longer on average with different odors?

```
                                             Individual 95% CIs For Mean
                                             Based on Pooled StDev
Level         n       Mean       StDev     -----+---------+---------+---------+-
lavender     29     106.07       13.18                          (------*-----)
lemon        28      89.79       15.44     (------*------)
none         30      91.27       14.93       (------*------)
                                           -----+---------+---------+---------+-
Pooled StDev =       14.54                   88.0      96.0     104.0     112.0
```

Is the condition of equal population standard deviation met? Why or why not?

a. No, because standard errors are given, not standard deviations.

b. No, because the sample standard deviations are different.

c. Yes, because the largest sample standard deviation is not more than twice the smallest one.

2 [Do good smells bring good business?](#) Do customers stay longer on average with different odors?

```
                                         Individual 95% CIs For Mean
                                         Based on Pooled StDev
Level       n      Mean     StDev   -----+---------+---------+---------+-
lavender   29    106.07     13.18                       (------*-----)
lemon      28     89.79     15.44   (------*------)
none       30     91.27     14.93     (------*------)
                                    -----+---------+---------+---------+-
Pooled StDev =    14.54                 88.0      96.0     104.0     112.0
```

Is the condition of equal population standard deviation met? Why or why not?

a. No, because standard errors are given, not standard deviations.

b. No, because the sample standard deviations are different.

c. Yes, because the largest sample standard deviation is not more than twice the smallest one.

2 **Do good smells bring good business?** A study was conducted at a small pizza restaurant in France on Saturday evenings in May. On one evening a relaxing lavender odor was spread through the restaurant; on another, a stimulating lemon odor; and a third evening had no odor. Treatments were randomly assigned to days. On each evening the time (in minutes) customers stayed in the restaurant was recorded.

Is the condition of randomization met? Why or why not?

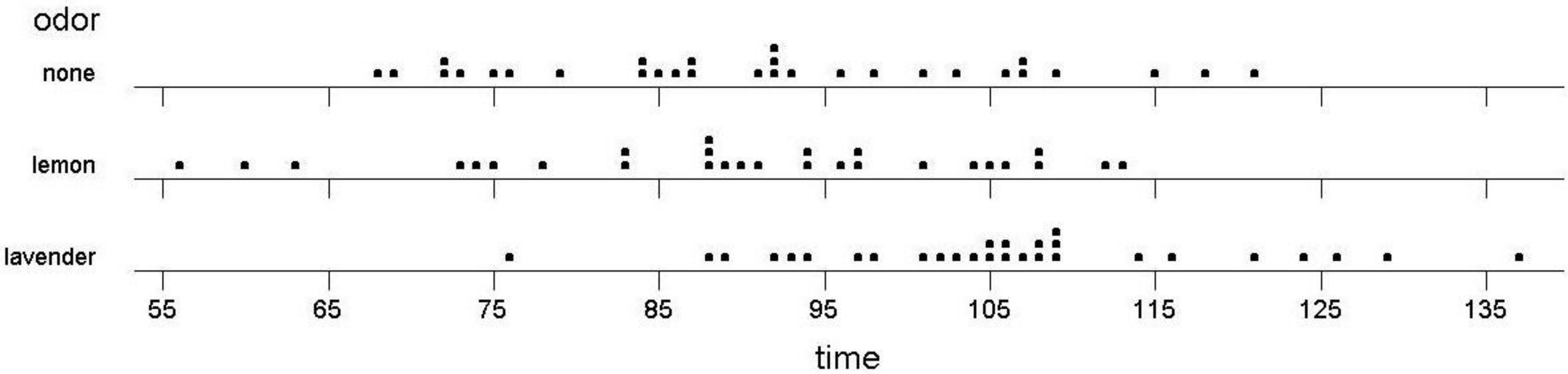
a. No, because odors were not randomly assigned to days.
b. No, customers were not randomly selected.
c. Yes, because the odors were randomly assigned to the days.

2 **Do good smells bring good business?** A study was conducted at a small pizza restaurant in France on Saturday evenings in May. On one evening a relaxing lavender odor was spread through the restaurant; on another, a stimulating lemon odor; and a third evening had no odor. Treatments were randomly assigned to days. On each evening the time (in minutes) customers stayed in the restaurant was recorded.

Is the condition of randomization met? Why or why not?

a. No, because odors were not randomly assigned to days.
b. No, customers were not randomly selected.
c. Yes, because the odors were randomly assigned to the days.

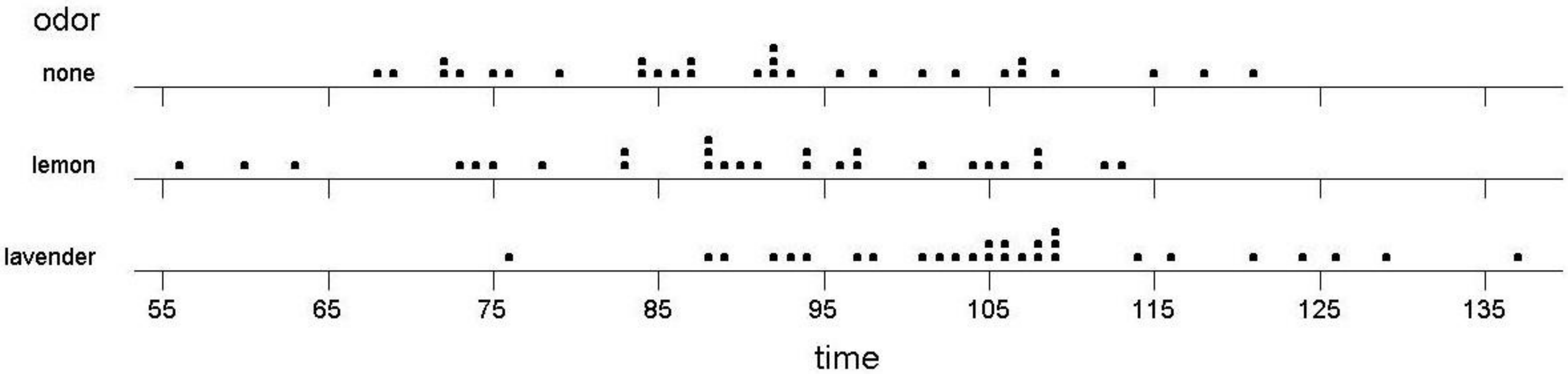
2 **Do good smells bring good business?** Do customers stay longer on average with different odors?



Is the condition of Normality met? Why or why not?

a. No, the dotplots are not bell-shaped symmetric.
b. No, there are outliers in the data.
c. Yes, because there are no extreme outliers in the data.

2 [Do good smells bring good business?](#) Do customers stay longer on average with different odors?



Is the condition of Normality met? Why or why not?

a. No, the dotplots are not bell-shaped symmetric.
b. No, there are outliers in the data.
c. Yes, because there are no extreme outliers in the data.

2 **Do good smells bring good business?** Do customers stay longer on average with different odors?

Source	df	SS	MS	F-Stat	P-value
Odor type	2	4687	2343	11.08	0.000
Error	84	17760	211		
Total	86	22447			

Using the ANOVA printout, what is the *P*-value for testing equality of means?

a. 0.0006
b. 0.000
c. 0.1108

2 <u>**Do good smells bring good business?**</u> Do customers stay longer on average with different odors?

Source	**df**	**SS**	**MS**	**F-Stat**	**P-value**
Odor type	2	4687	2343	11.08	0.000
Error	84	17760	211		
Total	86	22447			

Using the ANOVA printout, what is the *P*-value for testing equality of means?

a. 0.0006
b. 0.000
c. 0.1108

2 **Do good smells bring good business?** Do customers stay longer on average with different odors?

Source	df	SS	MS	F-Stat	P-value
Odor type	2	4687	2343	11.08	0.000
Error	84	17760	211		
Total	86	22447			

On the basis of the *P*-value = 0.000, what can you conclude?

a. Fail to reject H_0 and conclude that the means do not differ significantly.

b. Fail to reject H_0 and conclude that the means differ significantly.

c. Reject H_0 and conclude that at least one mean differs from the others.

2 **Do good smells bring good business?** Do customers stay longer on average with different odors?

Source	df	SS	MS	F-Stat	P-value
Odor type	2	4687	2343	11.08	0.000
Error	84	17760	211		
Total	86	22447			

On the basis of the *P*-value = 0.000, what can you conclude?

a. Fail to reject H_0 and conclude that the means do not differ significantly.

b. Fail to reject H_0 and conclude that the means differ significantly.

c. Reject H_0 and conclude that at least one mean differs from the others.

2 <u>Do good smells bring good business?</u> Do customers stay longer on average with different odors?

```
                                         Individual 95% CIs For Mean
                                         Based on Pooled StDev
Level        n      Mean     StDev    -----+---------+---------+---------+-
lavender    29    106.07     13.18                         (------*-----)
lemon       28     89.79     15.44    (------*-------)
none        30     91.27     14.93      (------*-------)
                                      -----+---------+---------+---------+-
Pooled StDev =     14.54                   88.0      96.0     104.0     112.0
```

On the basis of the confidence intervals, what can you conclude?

a. The mean time for the lemon odor is significantly greater than the other two means.
b. The mean time for lavender odor is significantly greater than the other two means.
c. The mean time for lemon and lavender odors is significantly greater than no odor.

2 **Do good smells bring good business?** Do customers stay longer on average with different odors?

```
                                          Individual 95% CIs For Mean
                                          Based on Pooled StDev
Level       n      Mean     StDev  -----+---------+---------+---------+-
lavender   29    106.07     13.18                      (------*-----)
lemon      28     89.79     15.44  (------*------)
none       30     91.27     14.93    (------*------)
                                   -----+---------+---------+---------+-
Pooled StDev =    14.54                88.0      96.0     104.0     112.0
```

On the basis of the confidence intervals, what can you conclude?

a. The mean time for the lemon odor is significantly greater than the other two means.

b. The mean time for lavender odor is significantly greater than the other two means.

c. The mean time for lemon and lavender odors is significantly greater than no odor.

3 A sample survey by the Pew Internet and American Life Project asked a random sample of adults about use of the Internet and about the type of community they lived in. Is there are relationship between type of community and use of the Internet? Here are the data in this two-way table:

	Community Type		
	Rural	Suburban	Urban
Internet users	433	1072	536
Nonusers	463	627	388

3 **23.29 Who's online?** A sample survey by the Pew Internet and American Life Project asked a random sample of adults about use of the Internet and about the type of community they lived in. Is there are relationship between type of community and use of the Internet?

Is this an experiment or an observational study? Can we conclude causation?

a. experiment—can conclude causation
b. experiment—cannot conclude causation
c. observational study—can conclude causation
d. observational study—cannot conclude causation

3 **23.29 Who's online?** A sample survey by the Pew Internet and American Life Project asked a random sample of adults about use of the Internet and about the type of community they lived in. Is there are relationship between type of community and use of the Internet?

Is this an experiment or an observational study? Can we conclude causation?

a. experiment—can conclude causation
b. experiment—cannot conclude causation
c. observational study—can conclude causation
d. observational study—cannot conclude causation

3 23.29 Who's online? A sample survey by the Pew Internet and American Life Project asked a random sample of adults about use of the Internet and about the type of community they lived in. Is there are relationship between type of community and use of the Internet?

What type of sampling design was used?

a. simple random sample
b. stratified
c. multistage
d. cluster

3 **23.29 Who's online?** A sample survey by the Pew Internet and American Life Project asked a random sample of adults about use of the Internet and about the type of community they lived in. Is there are relationship between type of community and use of the Internet?

What type of sampling design was used?

a. simple random sample
b. stratified
c. multistage
d. cluster

3 **23.29 Who's online?** A sample survey by the Pew Internet and American Life Project asked a random sample of adults about use of the Internet and about the type of community they lived in. Is there are relationship between type of community and use of the Internet?

What is the response variable? What type of variable?

a. Internet usage—categorical
b. Internet usage—quantitative
c. Type of community—categorical
d. Type of community—quantitative

3 **23.29 Who's online?** A sample survey by the Pew Internet and American Life Project asked a random sample of adults about use of the Internet and about the type of community they lived in. Is there are relationship between type of community and use of the Internet?

What is the response variable? What type of variable?

a. Internet usage—categorical
b. Internet usage—quantitative
c. Type of community—categorical
d. Type of community—quantitative

3 **23.29 Who's online?** A sample survey by the Pew Internet and American Life Project asked a random sample of adults about use of the Internet and about the type of community they lived in. Is there are relationship between type of community and use of the Internet?

What is the explanatory variable? What type of variable?

a. Internet usage—categorical
b. Internet usage—quantitative
c. Type of community—categorical
d. Type of community—quantitative

3 **23.29 Who's online?** A sample survey by the Pew Internet and American Life Project asked a random sample of adults about use of the Internet and about the type of community they lived in. Is there are relationship between type of community and use of the Internet?

What is the explanatory variable? What type of variable?

a. Internet usage—categorical
b. Internet usage—quantitative
c. Type of community—categorical
d. Type of community—quantitative

3 **23.29 Who's online?** A sample survey by the Pew Internet and American Life Project asked a random sample of adults about use of the Internet and about the type of community they lived in. Is there are relationship between type of community and use of the Internet?

What statistical procedure should be performed to answer the research question?

a. two-sample mean test
b. chi-square test
c. ANOVA
d. two-sample z proportion test

3 **23.29 Who's online?** A sample survey by the Pew Internet and American Life Project asked a random sample of adults about use of the Internet and about the type of community they lived in. Is there are relationship between type of community and use of the Internet?

What statistical procedure should be performed to answer the research question?

a. two-sample mean test
b. chi-square test
c. ANOVA
d. two-sample z proportion test

3 **23.29 Who's online?** A sample survey by the Pew Internet and American Life Project asked a random sample of adults about use of the Internet and about the type of community they lived in. Is there are relationship between type of community and use of the Internet?

State the null hypothesis for testing independence.

a. H_0: $p_1 = p_2 = p_3$
b. H_0: No relationship between type of community and internet usage.
c. H_0: There is a positive association between type of community and internet usage.

3 **23.29 Who's online?** A sample survey by the Pew Internet and American Life Project asked a random sample of adults about use of the Internet and about the type of community they lived in. Is there are relationship between type of community and use of the Internet?

State the null hypothesis for testing independence.

a. H_0: $p_1 = p_2 = p_3$ (conditional distributions are the same for all 3 communities)
b. H_0: No relationship between type of community and internet usage.
c. H_0: There is a positive association between type of community and internet usage.

Since these data are from an SRS with two questions, we should do a test of independence. Answer "a" would be valid if these data were from a stratified sample of community type.

3 23.29 Who's online? A sample survey by the Pew Internet and American Life Project asked a random sample of adults about use of the Internet and about the type of community they lived in. Is there are relationship between type of community and use of the Internet?

State the alternative hypothesis for testing independence.

a. H_a: At least one p_i differs from the others.

b. H_a: There is an association between type of community and internet usage.

c. H_a: No association between type of community and internet usage.

3 **23.29 Who's online?** A sample survey by the Pew Internet and American Life Project asked a random sample of adults about use of the Internet and about the type of community they lived in. Is there are relationship between type of community and use of the Internet?

State the alternative hypothesis for testing independence.

a. H_a: At least one p_i differs from the others.

b. H_a: There is an association between type of community and internet usage.

c. H_a: No association between type of community and internet usage.

3 **23.29 Who's online?** Is there are relationship between type of community and use of the Internet?

	Community Type			
	Rural	Suburban	Urban	Total
Internet users	433	1072	536	2041
Nonusers	463	627	388	1478
Total	896	1699	924	3519

How many Rural adults do we expect to be Internet users?

a. (433)(2041)/896
b. (433)(2041)(896)/3519
c. (2041)(1699)/3519
d. (2041)(896)/3519

3 **23.29 Who's online?** Is there are relationship between type of community and use of the Internet?

	Community Type			
	Rural	Suburban	Urban	Total
Internet users	433	1072	536	2041
Nonusers	463	627	388	1478
Total	896	1699	924	3519

How many Rural adults do we expect to be Internet users?

a. (433)(2041)/896
b. (433)(2041)(896)/3519
c. (2041)(1699)/3519
d. (2041)(896)/3519

3 Is the condition on expected counts met? Why or why not?

	Rural	Suburban	Urban	Total
Internet	433 (21.22%) 519.7	1072 (52.52%) 985.4	536 (26.26%) 535.9	2041 (100.00%)
No internet	463 (31.33%) 376.3	627 (42.42%) 713.6	388 (26.25%) 388.1	1478 (100.00%)
Total	896 (25.46%)	1699 (48.28%)	924 (26.26%)	3519 (100.00%)

a. Yes, because n = 3519 > 40
b. Yes, because all expected counts are > 5.
c. No, because the expected counts differ from the observed counts.

3 Is the condition on expected counts met? Why or why not?

	Rural	Suburban	Urban	Total
Internet	433 (21.22%) 519.7	1072 (52.52%) 985.4	536 (26.26%) 535.9	2041 (100.00%)
No internet	463 (31.33%) 376.3	627 (42.42%) 713.6	388 (26.25%) 388.1	1478 (100.00%)
Total	896 (25.46%)	1699 (48.28%)	924 (26.26%)	3519 (100.00%)

a. Yes, because n = 3519 > 40
b. Yes, because all expected counts are > 5.
c. No, because the expected counts differ from the observed counts.

3 What are the degrees of freedom?

	Rural	Suburban	Urban	Total
Internet	433 (21.22%) 519.7	1072 (52.52%) 985.4	536 (26.26%) 535.9	2041 (100.00%)
No internet	463 (31.33%) 376.3	627 (42.42%) 713.6	388 (26.25%) 388.1	1478 (100.00%)
Total	896 (25.46%)	1699 (48.28%)	924 (26.26%)	3519 (100.00%)

a. 2
b. 3
c. 4
d. 6
e. 3518

3 What are the degrees of freedom?

	Rural	Suburban	Urban	Total
Internet	433 (21.22%) 519.7	1072 (52.52%) 985.4	536 (26.26%) 535.9	2041 (100.00%)
No internet	463 (31.33%) 376.3	627 (42.42%) 713.6	388 (26.25%) 388.1	1478 (100.00%)
Total	896 (25.46%)	1699 (48.28%)	924 (26.26%)	3519 (100.00%)

a. 2 = (r – 1)(c – 1) = (2 – 1)(3 – 1)
b. 3
c. 4
d. 6
e. 3518

3 **23.29 Who's online?** Is there are relationship between type of community and use of the Internet?

Statistic	**DF**	**Value**	**P-value**
Chi-square	??	52.535	<0.0001

What are the values of the X^2 test statistic and *P*-value for testing independence?

3 **23.29 Who's online?** Is there are relationship between type of community and use of the Internet?

Statistic	DF	Value	P-value
Chi-square	??	52.535	<0.0001

What are the values of the X^2 test statistic and *P*-value for testing independence?

52.535 and $P < 0.0001$

3 **23.29 Who's online?** Is there are relationship between type of community and use of the Internet?

Statistic	**DF**	**Value**	**P-value**
Chi-square	??	52.535	<0.0001

On the basis of the *P*-value, what can you conclude?

a. There is a significant relationship between type of community and Internet usage.
b. There is insufficient evidence to conclude that there is a relationship between type of community and Internet usage.
c. Type of community causes some to use the Internet more than others.

3 [23.29 Who's online?] Is there are relationship between type of community and use of the Internet?

Statistic	DF	Value	P-value
Chi-square	??	52.535	<0.0001

On the basis of the *P*-value, what can you conclude?

a. There is a significant relationship between type of community and Internet usage.
b. There is insufficient evidence to conclude that there is a relationship between type of community and Internet usage.
c. Type of community causes some to use the Internet more than others.

True/false questions

9. A boxplot can be used to display both categorical and quantitative data.

    a. True
    b. False

True/false questions

9. A boxplot can be used to display both categorical and quantitative data.

   a. True
   b. False

False, boxplots are only used to display quantitative data.

Wabash Tech has two professional schools, business and law. The two-way tables to both schools are categorized by gender and admissions decision. The percentages in parentheses are the conditional distributions for the rows.

Business	Admit	Deny	TOTAL		**Law**	Admit	Deny	TOTAL
Male	480 (80%)	120 (20%)	600 (100%)		Male	10 (10%)	90 (90%)	100 (100%)
Female	180 (90%)	20 (10%)	200 (100%)		Female	100 (33.3%)	200 (66.7%)	300 (100%)
TOTAL	660 (82.5%)	140 (17.5%)	800 (100%)		TOTAL	110 (27.5%)	290 (72.5%)	400 (100%)

Overall	Admit	Deny	TOTAL
Male	490 (70%)	210 (30%)	700 (100%)
Female	280 (56%)	220 (44%)	500 (100%)
TOTAL	770 (64.2%)	430 (35.8%)	1200 (100%)

Business School versus Law School

b) For the combined table, what percent of the students are admitted?

Overall	Admit	Deny	TOTAL
Male	490 (70%)	210 (30%)	700 (100%)
Female	280 (56%)	220 (44%)	500 (100%)
TOTAL	770 (64.2%)	430 (35.8%)	1200 (100%)

a. 490/770
b. 490/1200
c. 280/770
d. 280/1200
e. 770/1200

Business School versus Law School

b) For the combined table, what percent of the students are admitted?

Overall	Admit	Deny	TOTAL
Male	490 (70%)	210 (30%)	700 (100%)
Female	280 (56%)	220 (44%)	500 (100%)
TOTAL	770 (64.2%)	430 (35.8%)	1200 (100%)

a. 490/770
b. 490/1200
c. 280/770
d. 280/1200
e. 770/1200

Correct answer is “e”

Business School versus Law School

c) For the combined table, what percent of the males are admitted?

Overall	Admit	Deny	TOTAL
Male	490 (70%)	210 (30%)	700 (100%)
Female	280 (56%)	220 (44%)	500 (100%)
TOTAL	770 (64.2%)	430 (35.8%)	1200 (100%)

a. 490/700
b. 490/770
c. 490/1200
d. 770/1200

Business School versus Law School

c) For the combined table, what percent of the males are admitted?

Overall	Admit	Deny	TOTAL
Male	490 (70%)	210 (30%)	700 (100%)
Female	280 (56%)	220 (44%)	500 (100%)
TOTAL	770 (64.2%)	430 (35.8%)	1200 (100%)

a. 490/700
b. 490/770
c. 490/1200
d. 770/1200

Correct answer is "a"

Business School versus Law School

d) For the combined table, what is the conditional distribution for gender given that they were admitted?

Overall	Admit	Deny	TOTAL
Male	490 (70%)	210 (30%)	700 (100%)
Female	280 (56%)	220 (44%)	500 (100%)
TOTAL	770 (64.2%)	430 (35.8%)	1200 (100%)

a. 490/700, 280/500
b. 490/770, 280/770
c. 490/1200, 280/1200
d. 770/1200, 430/1200

Business School versus Law School

d) For the combined table, what is the conditional distribution for gender given that they were admitted?

Overall	Admit	Deny	TOTAL
Male	490 (70%)	210 (30%)	700 (100%)
Female	280 (56%)	220 (44%)	500 (100%)
TOTAL	770 (64.2%)	430 (35.8%)	1200 (100%)

a. 490/700, 280/500
b. 490/770, 280/770
c. 490/1200, 280/1200
d. 770/1200, 430/1200

Correct answer is “b” with denominator = total number admitted

Business School versus Law School

e) For the combined table, what is the conditional distribution for admission decision for males?

Overall	Admit	Deny	TOTAL
Male	490 (70%)	210 (30%)	700 (100%)
Female	280 (56%)	220 (44%)	500 (100%)
TOTAL	770 (64.2%)	430 (35.8%)	1200 (100%)

a. 490/700, 210/700
b. 490/770, 210/430
c. 490/1200, 210/1200
d. 770/1200, 430/1200

Business School versus Law School

e) For the combined table, what is the conditional distribution for admission decision for males?

Overall	Admit	Deny	TOTAL
Male	490 (70%)	210 (30%)	700 (100%)
Female	280 (56%)	220 (44%)	500 (100%)
TOTAL	770 (64.2%)	430 (35.8%)	1200 (100%)

a. 490/700, 210/700
b. 490/770, 210/430
c. 490/1200, 210/1200
d. 770/1200, 430/1200

Correct answer is “a”, where both denominators are 700 = # of males

Is there an association between gender and admission decision?

Business School versus Law School

For the combined table, what hypotheses would you use to test whether there is a difference in the proportion being admitted for males and females?

Overall	Admit	Deny	TOTAL
Male	490 (70%)	210 (30%)	700 (100%)
Female	280 (56%)	220 (44%)	500 (100%)
TOTAL	770 (64.2%)	430 (35.8%)	1200 (100%)

a. $H_o: p_{male} = p_{female}$ vs. $H_a: p_{male} \neq p_{female}$
b. $H_o: p_{male} = p_{female}$ vs. $H_a: p_{male} > p_{female}$
c. $H_o: p_{male} = p_{female}$ vs. $H_a: p_{male} < p_{female}$
d. $H_o: \hat{p}_{male} = \hat{p}_{female}$ vs. $H_a: \hat{p}_{male} \neq \hat{p}_{female}$

Business School versus Law School

For the combined table, what hypotheses would you use to test whether there is a difference in the proportion being admitted for males and females?

Overall	Admit	Deny	TOTAL
Male	490 (70%)	210 (30%)	700 (100%)
Female	280 (56%)	220 (44%)	500 (100%)
TOTAL	770 (64.2%)	430 (35.8%)	1200 (100%)

a. $H_o: p_{male} = p_{female}$ vs. $H_a: p_{male} \neq p_{female}$
b. $H_o: p_{male} = p_{female}$ vs. $H_a: p_{male} > p_{female}$
c. $H_o: p_{male} = p_{female}$ vs. $H_a: p_{male} < p_{female}$
d. $H_o: \hat{p}_{male} = \hat{p}_{female}$ vs. $H_a: \hat{p}_{male} \neq \hat{p}_{female}$

Business School versus Law School

For the combined table, if the *P*-value < 0.0001, what conclusion should be made using $\alpha = 0.05$?

Overall	Admit	Deny	TOTAL
Male	490 (70%)	210 (30%)	700 (100%)
Female	280 (56%)	220 (44%)	500 (100%)
TOTAL	770 (64.2%)	430 (35.8%)	1200 (100%)

a. There is sufficient evidence to conclude that admittance rates are the same for males and females.
b. There is sufficient evidence to conclude that admittance rates are different for males and females.
c. There is insufficient evidence to conclude that admittance rates are the same for males and females.
d. There is insufficient evidence to conclude that admittance rates are different for males and females.

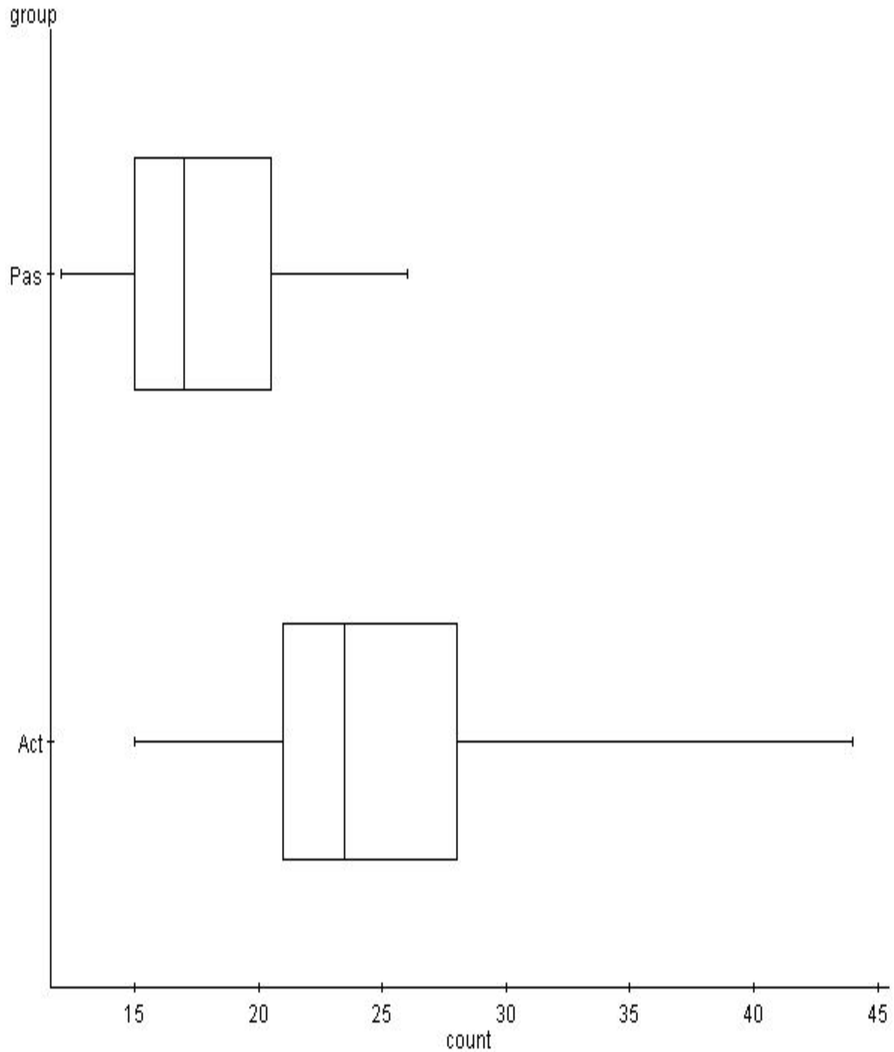
Business School versus Law School

For the combined table, if the *P*-value < 0.0001, what conclusion should be made using $\alpha = 0.05$?

Overall	Admit	Deny	TOTAL
Male	490 (70%)	210 (30%)	700 (100%)
Female	280 (56%)	220 (44%)	500 (100%)
TOTAL	770 (64.2%)	430 (35.8%)	1200 (100%)

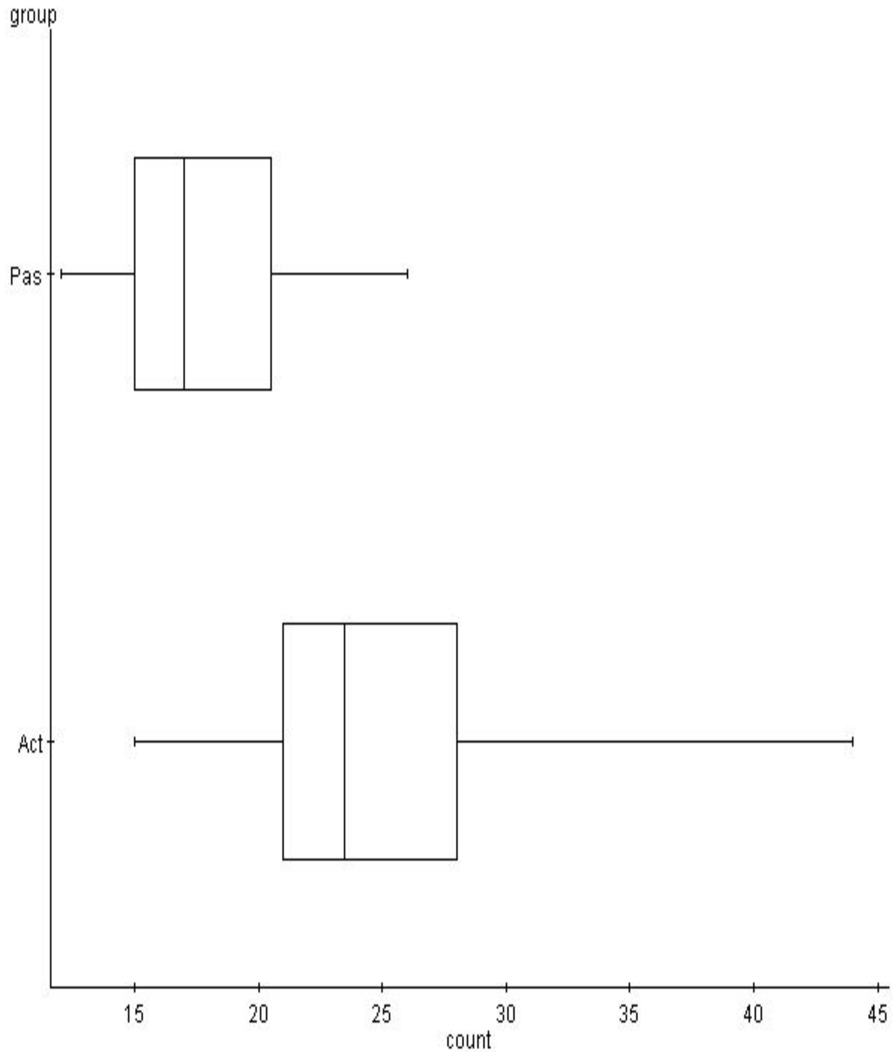
a. There is sufficient evidence to conclude that admittance rates are the same for males and females.
b. There is sufficient evidence to conclude that admittance rates are different for males and females.
c. There is insufficient evidence to conclude that admittance rates are the same for males and females.
d. There is insufficient evidence to conclude that admittance rates are different for males and females.

Here are the results of the test for two groups who studied the Blissymbols. One used passive learning and the other used active learning. How do the two boxplots compare?



Shape
Center
Spread

Here are the results of the test for two groups who studied the Blissymbols. One used passive learning and the other used active learning. How do the means compare with the medians?



Here are the results of the test for two groups who studied the Blissymbols. One used passive learning and the other used active learning. How do the two stemplots compare?

Count for passive		Count for Active
2234	1	
555566777889	1	567
00111	2	0011112334444
566	2	78889
	3	1
	3	5
	4	4

Here are the results of the test for two groups who studied the Blissymbols. One used passive learning and the other used active learning. What are the correct hypotheses for testing whether the mean score for the Active group is higher than the mean score for the passive group?

Count for passive		**Count for Active**
2234	1	
555566777889	1	567
00111	2	0011112334444
566	2	78889
	3	1
	3	5
	4	4

a. H_0: $\mu_{passive} = \mu_{active}$ vs. H_a: $\mu_{passive} \neq \mu_{active}$

b. H_0: $\mu_{passive} = \mu_{active}$ vs. H_a: $\mu_{passive} > \mu_{active}$

c. H_0: $\bar{x}_{male} = \bar{x}_{female}$ vs. H_a: $\bar{x}_{male} \neq \bar{x}_{active}$

d. H_0: $\mu_{passive} = \mu_{active}$ vs. H_a: $\mu_{passive} < \mu_{active}$

Here are the results of the test for two groups who studied the Blissymbols. One used passive learning and the other used active learning. What are the correct hypotheses for testing whether the mean score for the Active group is higher than the mean score for the passive group?

Count for passive		**Count for Active**
2234	1	
555566777889	1	567
00111	2	0011112334444
566	2	78889
	3	1
	3	5
	4	4

a. $H_0: \mu_{passive} = \mu_{active}$ vs. $H_a: \mu_{passive} \neq \mu_{active}$

b. $H_0: \mu_{passive} = \mu_{active}$ vs. $H_a: \mu_{passive} > \mu_{active}$

c. $H_0: \bar{x}_{male} = \bar{x}_{female}$ vs. $H_a: \bar{x}_{male} \neq \bar{x}_{active}$

d. $H_0: \mu_{passive} = \mu_{active}$ vs. $H_a: \mu_{passive} < \mu_{active}$

Here are the results of the test for two groups who studied the Blissymbols. One used passive learning and the other used active learning. If the *P*-value = 0.00005 and α = 0.05, what would you conclude?

Count for passive		Count for Active
2234	1	
555566777889	1	567
00111	2	0011112334444
566	2	78889
	3	1
	3	5
	4	4

H_0: $\mu_{passive} = \mu_{active}$

H_a: $\mu_{passive} < \mu_{active}$

a. The mean scores for the Passive and Active groups are not significantly different.
b. The mean score is significantly lower for the Passive group.
c. The mean score is significantly higher for the Passive group.
d. The mean scores for the Passive and Active groups are not signficantly different.

Here are the results of the test for two groups who studied the Blissymbols. One used passive learning and the other used active learning. If the *P*-value = 0.00005 and $\alpha = 0.05$, what would you conclude?

Count for passive		**Count for Active**
2234	1	
555566777889	1	567
00111	2	0011112334444
566	2	78889
	3	1
	3	5
	4	4

$H_0: \mu_{passive} = \mu_{active}$

$H_a: \mu_{passive} < \mu_{active}$

a. The mean scores for the Passive and Active groups are not significantly different.
b. The mean score is significantly lower for the Passive group.
c. The mean score is significantly higher for the Passive group.
d. The mean scores for the Passive and Active groups are not signficantly different.

Are you a morning person, an evening person, or neither? A sample of 100 students were randomly selected from the student body of a small private university. Each took a psychological test that found 16 morning, 30 evening and 54 who were neither.

What is the response variable?
Is it categorical or quantitative?
What type of graph is appropriate for these data?

Are you a morning person, an evening person, or neither? A sample of 100 students were randomly selected from the student body of a small private university. Each took a psychological test that found 16 morning, 30 evening and 54 who were neither.

What is the response variable?
Is it categorical or quantitative?
What type of graph is appropriate for these data?

Response variable is whether student is a morning person, an evening person or neither. It is categorical. Appropriate graph is a bar graph.

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: “Do cars get better miles per gallon on average with clean air filters?”

Cars	1	2	3	4	5	6	7	8	9	10
Clean--mpg	19	22	24	24.5	25	25	25.5	26	28	31
Dirty--mpg	16	20	21	21.5	23	21	22.5	25	25	27
Differences	3	2	3	3	2	4	3	1	3	4

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: “Do cars get better miles per gallon on average with clean air filters?”

   A. What type of study is this?
      a. Observational study
      b. Randomized controlled experiment
      c. Matched pairs experiment

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: “Do cars get better miles per gallon on average with clean air filters?”

   A. What type of study is this?
      a. Observational study
      b. Randomized controlled experiment
      c. Matched pairs experiment

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: “Do cars get better miles per gallon on average with clean air filters?”

B. Do we need a test of significance or a confidence interval?

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: "Do cars get better miles per gallon on average with clean air filters?"

B. Do we need a test of significance or a confidence interval?

Test of significance

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: “Do cars get better miles per gallon on average with clean air filters?”

C. What procedure should be used to answer the question?

a. One-sample *t*
b. Matched-pairs *t*

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: “Do cars get better miles per gallon on average with clean air filters?”

C. What procedure should be used to answer the question?

a. One-sample *t*
b. Matched-pairs *t*: two measurements on each car

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: “Do cars get better miles per gallon on average with clean air filters?”

D. What are the appropriate hypotheses?

E. Is this test one-sided or two-sided?

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: "Do cars get better miles per gallon on average with clean air filters?"

D. What are the appropriate hypotheses?

H_0: $\mu_d = 0$ versus H_a: $\mu_d > 0$

E. Is this test one-sided or two-sided?

One-sided test

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: "Do cars get better miles per gallon on average with clean air filters?"

D. What are the appropriate hypotheses?

H_0: $\mu_d = 0$ versus H_a: $\mu_d > 0$

What is the parameter being tested in context?

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: “Do cars get better miles per gallon on average with clean air filters?”

D. What are the appropriate hypotheses?

H_0: $\mu_d = 0$ versus H_a: $\mu_d > 0$

What is the parameter being tested in context?

μ_d = mean difference between mpg with clean air filters and mpg with dirty air filters

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: “Do cars get better miles per gallon on average with clean air filters?”

E. What conditions should be checked?

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: "Do cars get better miles per gallon on average with clean air filters?"

E. What conditions should be checked?

Randomization (data collection) and Normality of the population of differences.

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: “Do cars get better miles per gallon on average with clean air filters?”

E. What conditions should be checked?

Randomization (data collection) and Normality of the population of differences.

Is the randomization condition met?

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: "Do cars get better miles per gallon on average with clean air filters?"

E. What conditions should be checked?

Randomization (data collection) and Normality of the population of differences.

Is the randomization condition met?

Yes, because the order of testing was randomized.

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: “Do cars get better miles per gallon on average with clean air filters?”

F. For the Normality condition, which graph or graphs need to be checked for outliers?

Stem and leaf of **clean** mpg
N = 10
Leaf Unit = 1.0
1|9
2|
2|2
2|44555
2|6
2|8
3|1

Stem and leaf of **dirty** mpg
N = 10
Leaf Unit = 1.0
1|6
1|
2|0111
2|23
2|55
2|7

Stem and leaf of **differences**
N = 10
Leaf Unit = 0.10
1|0
2|00
3|00000
4|00

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: “Do cars get better miles per gallon on average with clean air filters?”

F. For the Normality condition, which graph or graphs need to be checked for outliers?

Stem and leaf of **clean** mpg
N = 10
Leaf Unit = 1.0

1|9
2|
2|2
2|44555
2|6
2|8
3|1

Stem and leaf of **dirty** mpg
N = 10
Leaf Unit = 1.0

1|6
1|
2|0111
2|23
2|55
2|7

Stem and leaf of **differences**
N = 10
Leaf Unit = 0.10

1|0
2|00
3|00000
4|00

Must check a graph of the differences

G. What is the value of the standardized test statistic? What should be used for μ_0?

Variable	N	Mean	Median	Tr Mean	StDev	SE Mean
clean	10	25.00	25.00	25.00	3.21	1.01
dirty	10	22.200	22.000	2.375	3.093	0.978
diff	10	2.800	3.000	2.875	0.919	0.291

G. What is the value of the standardized test statistic? What should be used for μ_0?

Variable	N	Mean	Median	Tr Mean	StDev	SE Mean
clean	10	25.00	25.00	25.00	3.21	1.01
dirty	10	22.200	22.000	2.375	3.093	0.978
diff	10	2.800	3.000	2.875	0.919	0.291

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} = \frac{2.8 - 0}{0.919 / \sqrt{10}} = 9.63$$

μ_0 = zero

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: “Do cars get better miles per gallon on average with clean air filters?”

H. What are the degrees of freedom?

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: “Do cars get better miles per gallon on average with clean air filters?”

H. What are the degrees of freedom?

df = 10 – 1 = 9 where # of pairs = 10

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: “Do cars get better miles per gallon on average with clean air filters?”

H. What are the degrees of freedom?

df = 10 – 1 = 9   where # of pairs = 10

What is the *P*-value from the *t* table?

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: "Do cars get better miles per gallon on average with clean air filters?"

H. What are the degrees of freedom?

df = 10 – 1 = 9 where # of pairs = 10

What is the *P*-value from the *t* table?

Look on df = 9 row and find *t* = 9.63 off the chart

So, *P*-value < 0.0005.

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: “Do cars get better miles per gallon on average with clean air filters?”

I. For *P*-value = 0.000 and α = 0.05, are these results statistically significant?

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: "Do cars get better miles per gallon on average with clean air filters?"

I. For P-value = 0.000 and $\alpha = 0.05$, are these results statistically significant?

Yes, because P-value $= 0.000 < 0.05 = \alpha$.

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: "Do cars get better miles per gallon on average with clean air filters?"

J. What conclusions should be made?

Reject H_0 and conclude that the cars get better miles per gallon on average with clean air filters.

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: "Do cars get better miles per gallon on average with clean air filters?"

K. Are the conclusions also practically important?

Why or why not?

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: "Do cars get better miles per gallon on average with clean air filters?"

K. Are the conclusions also practically important?

Yes

Why or why not?

Because a mean difference of 2.8 miles per gallon is definitely worth buying a clean air filter.

Note: $\bar{x}_d - \mu_0 = 2.8 - 0 = 2.8$.

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: "What is the mean difference in mpg for clean air filters and dirty air filters?"

Do we need a test of significance or a confidence interval?

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: "What is the mean difference in mpg for clean air filters and dirty air filters?"

Do we need a test of significance or a confidence interval?

Confidence Interval

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: "What is the mean difference in mpg for clean air filters and dirty air filters?"

Variable	N	Mean	Median	Tr Mean	StDev	SE Mean
clean	10	25.00	25.00	25.00	3.21	1.01
dirty	10	22.200	22.000	2.375	3.093	0.978
diff	10	2.800	3.000	2.875	0.919	0.291

Compute and interpret the confidence interval.

3. Gas mileage for 10 cars with dirty air filters and clean air filters was studied. Each car was tested once with a clean air filter and once with a dirty air filter (with the order of the testing randomized.) The research question is: "What is the mean difference in mpg for clean air filters and dirty air filters?"

Variable	N	Mean	Median	Tr Mean	StDev	SE Mean
clean	10	25.00	25.00	25.00	3.21	1.01
dirty	10	22.200	22.000	2.375	3.093	0.978
diff	10	2.800	3.000	2.875	0.919	0.291

Compute and interpret the confidence interval.

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}} = 2.8 \pm 2.262\left(\frac{0.919}{\sqrt{10}}\right) = (2.14, 3.46)$$

We are 95% confident that the true mean difference between mpg for clean air filters and dirty air filters is between 2.14 mpg and 3.46 mpg.

$\bar{x}$

General true / false questions

34. One way to have matched pairs is to have two identical measurements on each individual (e.g., each boy wears a shoe with sole material A on one foot and a shoe with sole material B on the other foot—with right or left randomized; each athlete tests both apparatus A and apparatus B with order randomized, etc.)
    a. True
    b. False

$\bar{x}$

General true / false questions

34. One way to have matched pairs is to have two identical measurements on each individual (e.g., each boy wears a shoe with sole material A on one foot and a shoe with sole material B on the other foot—with right or left randomized; each athlete tests both apparatus A and apparatus B with order randomized, etc.)

    a. True
    b. False

Two measurements on each individual is the most common form of matched pairs.

$\bar{x}$

General true / false questions

35. We always plot the differences when checking the conditions for a matched paired *t* test.

  a. True
  b. False

$\bar{x}$

General true / false questions

35. We always plot the differences when checking the conditions for a matched paired *t* test.

a. True
b. False

$\bar{x}$

General true / false questions

36. The *P*-value for testing H_0: $\mu_d = 0$ vs. H_a: $\mu_d \neq 0$ for $t = -0.78$ with df = 19 is 0.40 < P-value < 0.50.

a. True
b. False

$\bar{x}$

General true / false questions

36. The *P*-value for testing H_0: $\mu_d = 0$ vs. H_a: $\mu_d \neq 0$ for
$t = -0.78$ with df = 19 is 0.40 < P-value < 0.50.

a. True
b. False

General true / false questions

37. We can use a confidence interval approach to a test of significance if the alternative hypothesis is two-sided and the confidence level and significance level add up to 100%.
    a. True
    b. False

General true / false questions

37. We can use a confidence interval approach to a test of significance if the alternative hypothesis is two-sided and the confidence level and significance level add up to 100%.
    a. True
    b. False

Researchers at the National Cancer Institute released the results of a study that examined the effect of exposing dogs to weed-killing herbicides. Let

p_1 = percentage of unexposed dogs that developed lymphoma

p_2 = percentage of exposed dogs that developed lymphoma

The 90% confidence interval for $p_1 - p_2$ is (-0.484, -0.368). On the basis of this interval, what should we conclude?

A. There is no significant difference in the incidence rates of lymphoma between exposed dogs and unexposed dogs.

B. Dogs exposed to weed-killing herbicides have a higher incidence of lymphoma than unexposed dogs.

C. Dogs exposed to weed-killing herbicides have a lower incidence of lymphoma than unexposed dogs.

D. Since percentages cannot be negative, no conclusion can be made.

Researchers at the National Cancer Institute released the results of a study that examined the effect of exposing dogs to weed-killing herbicides. Let

p_1 = percentage of unexposed dogs that developed lymphoma

p_2 = percentage of exposed dogs that developed lymphoma

The 90% confidence interval for $p_1 - p_2$ is (-0.484, -0.368). On the basis of this interval, what should we conclude?

A. There is no significant difference in the incidence rates of lymphoma between exposed dogs and unexposed dogs.

B. Dogs exposed to weed-killing herbicides have a higher incidence of lymphoma than unexposed dogs.

C. Dogs exposed to weed-killing herbicides have a lower incidence of lymphoma than unexposed dogs.

D. Since percentages cannot be negative, no conclusion can be made.

The end