Stat 13, Intro. to Statistical Methods for the Life and Health Sciences.

1. Practice problems.

Read ch9.

http://www.stat.ucla.edu/~frederic/13/W25 .

The final is Wed Mar19, 1130am-230pm, here in Franz 1178, and
will be on ch 1-7, 10, and 1 question on ch9.
Bring a PENCIL or pen and CALCULATOR and any books or notes you want.
No computers.

If you cannot take it because of an emergency or other health reason,
then you will get an incomplete in the course and need to arrange to take the Spring stat 13 final.

Suppose that among a sample of 100 adults in a given town, the correlation between height (inches) and weight (lbs.) is 0.82, and the mean height is 65 inches, the median height is 64.5 inches, the sd of height is 5 inches, the mean weight is 160 lbs., and the sd of weight is 40 lbs.

5. Find an approximate 95% CI for the correlation between height and weight for the whole population.

a. 0.82 +/- 0.0225.

b. 0.82 +/- 0.0397.

c. 0.82 +/- 0.113.

d. 0.82 +/- 0.299.

e. 0.82 +/- 0.324.

Suppose that among a sample of 100 adults in a given town, the correlation between height (inches) and weight (lbs.) is 0.82, and the mean height is 65 inches, the median height is 64.5 inches, the sd of height is 5 inches, the mean weight is 160 lbs., and the sd of weight is 40 lbs.

5. Find an approximate 95% CI for the correlation between height and weight for the whole population.

a. 0.82 +/- 0.0225.

b. 0.82 +/- 0.0397.

c. 0.82 +/- 0.113.

d. 0.82 +/- 0.299.

e. 0.82 +/- 0.324.

1.96*sqrt((1-.82^2)/(100-2)) ~ 0.113.

(Technically since it is a t-CI, the multiplier should be based on the t-distribution and should be 1.98 instead of 1.96.)

Suppose that among a sample of 100 adults in a given town, the correlation between height (inches) and weight (lbs.) is 0.82, and the mean height is 65 inches, the median height is 64.5 inches, the sd of height is 5 inches, the mean weight is 160 lbs., and the sd of weight is 40 lbs.

6. Find an approximate 95% CI for the estimated slope for predicting weight using height, for the whole population.

a. 6.56 +/- 0.0978.

b. 6.56 +/- 0.124.

c. 6.56 +/- 0.235.

d. 6.56 +/- 0.337.

e. 6.56 +/- 0.463.

Suppose that among a sample of 100 adults in a given town, the correlation between height (inches) and weight (lbs.) is 0.82, and the mean height is 65 inches, the median height is 64.5 inches, the sd of height is 5 inches, the mean weight is 160 lbs., and the sd of weight is 40 lbs.

6. Find an approximate 95% CI for the estimated slope for predicting weight using height, for the whole population.

a. 6.56 +/- 0.0978.

b. 6.56 +/- 0.124.

c. 6.56 +/- 0.235.

d. 6.56 +/- 0.337.

e. 6.56 +/- 0.907.

r*sy/sx +/- 1.96*sqrt((1-r^2)/(n-2))*sy/sx = 6.56 +/- 0.907.

Suppose that among a sample of 100 adults in a given town, the correlation between height (inches) and weight (lbs.) is 0.82, and the mean height is 65 inches, the median height is 64.5 inches, the sd of height is 5 inches, the mean weight is 160 lbs., and the sd of weight is 40 lbs.

7. Why shouldn't one trust this regression line to predict the weight of someone who is 25 inches tall?

a. The sample size is insufficiently large.

b. The sample SD of weight is too small.

c. The value of 25 inches is too far outside the range of most observations.

d. The correlation of the ANOVA is a t-test confidence interval with statistical significance.

e. The data come from an observational study, so there may be confounding factors.

f. The height values are heavily right skewed, so the prediction errors are large.

Suppose that among a sample of 100 adults in a given town, the correlation between height (inches) and weight (lbs.) is 0.82, and the mean height is 65 inches, the median height is 64.5 inches, the sd of height is 5 inches, the mean weight is 160 lbs., and the sd of weight is 40 lbs.

7. Why shouldn't one trust this regression line to predict the weight of someone who is 25 inches tall?

a. The sample size is insufficiently large.

b. The sample SD of weight is too small.

c. The value of 25 inches is too far outside the range of most observations.

d. The correlation of the ANOVA is a t-test confidence interval with statistical significance.

e. The data come from an observational study, so there may be confounding factors.

f. The height values are heavily right skewed, so the prediction errors are large.

Suppose that among a sample of 100 adults in a given town, the correlation between height (inches) and weight (lbs.) is 0.82, and the mean height is 65 inches, the median height is 64.5 inches, the sd of height is 5 inches, the mean weight is 160 lbs., and the sd of weight is 40 lbs. The estimated slope in predicting weight from height is 6.56 lbs/in.

8. How should one interpret the estimated slope of 6.56?

a. Each extra inch you grow causes you to increase your weight by 6.56 lbs on average.

b. Each extra lb. you weigh causes you to grow 6.56 inches.

c. The amount of weight Americans average is 6.56 standard errors above the mean.

d. The Z-score corresponding to the correlation between height and weight is 6.56.

e. For each extra inch taller you are, your predicted weight increases by 6.56 lbs.

f. The proportion of variance in weight explained by the regression equation is 6.56%.

Suppose that among a sample of 100 adults in a given town, the correlation between height (inches) and weight (lbs.) is 0.82, and the mean height is 65 inches, the median height is 64.5 inches, the sd of height is 5 inches, the mean weight is 160 lbs., and the sd of weight is 40 lbs. The estimated slope in predicting weight from height is 6.56 lbs/in.

8. How should one interpret the estimated slope of 6.56?

a. Each extra inch you grow causes you to increase your weight by 6.56 lbs on average.

b. Each extra lb. you weigh causes you to grow 6.56 inches.

c. The amount of weight Americans average is 6.56 standard errors above the mean.

d. The Z-score corresponding to the correlation between height and weight is 6.56.

e. For each extra inch taller you are, your predicted weight increases by 6.56 lbs.

f. The proportion of variance in weight explained by the regression equation is 6.56%.

Suppose a researcher studying acne, or oily skin, takes a simple random sample of 400 Americans age 21-30 and calls them group A, and a simple random sample of 200 Americans age 31-40 and calls them group B. In group A, 88 people have acne, and in group B, 30 people have acne.

9. What is the pooled sample percentage with acne, in both groups combined?

a. 16.1%. b. 17.2%. c. 18.4%. d. 19.7%. e. 21.2%. f. 23.0%.

Suppose a researcher studying acne, or oily skin, takes a simple random sample of 400 Americans age 21-30 and calls them group A, and a simple random sample of 200 Americans age 31-40 and calls them group B. In group A, 88 people have acne, and in group B, 30 people have acne.

9. What is the pooled sample percentage with acne, in both groups combined?

a. 16.1%. b. 17.2%. c. 18.4%. **d. 19.7%.** e. 21.2%. f. 23.0%.

118/600 = 19.7%.

Suppose a researcher studying acne, or oily skin, takes a simple random sample of 400 Americans age 21-30 and calls them group A, and a simple random sample of 200 Americans age 31-40 and calls them group B. In group A, 88 people have acne, and in group B, 30 people have acne.

10. Using this pooled sample percentage, under the null hypothesis that the two groups have the same acne rate, what is the standard error for the difference between the two percentages?

a. 1.42%. b. 1.88%. c. 2.02%. d. 2.99%. e. 3.08%. f. 3.44%.

Suppose a researcher studying acne, or oily skin, takes a simple random sample of 400 Americans age 21-30 and calls them group A, and a simple random sample of 200 Americans age 31-40 and calls them group B. In group A, 88 people have acne, and in group B, 30 people have acne.

10. Using this pooled sample percentage, under the null hypothesis that the two groups have the same acne rate, what is the standard error for the difference between the two percentages?

a. 1.42%. b. 1.88%. c. 2.02%. d. 2.99%. e. 3.08%. **f. 3.44%.**

$\sqrt{(.197 * (1-.197)/400 + .197*(1-.197)/200)} = .0344$.

Suppose a researcher studying acne, or oily skin, takes a simple random sample of 400 Americans age 21-30 and calls them group A, and a simple random sample of 200 Americans age 31-40 and calls them group B. In group A, 88 people have acne, and in group B, 30 people have acne.

11. What is the Z statistic for the difference between the two group percentages?

a. 1.02. b. 1.23. c. 1.55. d. 1.88. e. 2.03. f. 2.43.

Suppose a researcher studying acne, or oily skin, takes a simple random sample of 400 Americans age 21-30 and calls them group A, and a simple random sample of 200 Americans age 31-40 and calls them group B. In group A, 88 people have acne, and in group B, 30 people have acne.

11. What is the Z statistic for the difference between the two group percentages?

a. 1.02. b. 1.23. c. 1.55. d. 1.88. **e. 2.03.** f. 2.43.

(88/400 – 30/200) ÷ 0.0344 = 2.03.

Suppose a researcher studying acne, or oily skin, takes a simple random sample of 400 Americans age 21-30 and calls them group A, and a simple random sample of 200 Americans age 31-40 and calls them group B. In group A, 88 people have acne, and in group B, 30 people have acne.

12. Using the unpooled standard error for the difference between the two percentages, find a 95% confidence interval for the percentage with acne among those age 21-30 minus the percentage of acne among those age 31-40.

a. 7% +/- 5.02%. b. 7% +/- 5.54%. c. 7% +/- 5.92%. d. 7% +/- 6.03%. e. 7% +/- 6.41%.

Suppose a researcher studying acne, or oily skin, takes a simple random sample of 400 Americans age 21-30 and calls them group A, and a simple random sample of 200 Americans age 31-40 and calls them group B. In group A, 88 people have acne, and in group B, 30 people have acne.

12. Using the unpooled standard error for the difference between the two percentages, find a 95% confidence interval for the percentage with acne among those age 21-30 minus the percentage of acne among those age 31-40.

a. 7% +/- 5.02%. b. 7% +/- 5.54%. c. 7% +/- 5.92%. d. 7% +/- 6.03%. **e. 7% +/- 6.41%.**

$88/400 - 30/200 \pm 1.96 * \sqrt{(88/400 * (1-88/400) / 400 + 30/200 * (1-30/200) / 200)}$

$= 7\% \pm 6.41\%$.

Suppose a researcher studying acne, or oily skin, takes a simple random sample of 400 Americans age 21-30 and calls them group A, and a simple random sample of 200 Americans age 31-40 and calls them group B. In group A, 88 people have acne, and in group B, 30 people have acne.

13. What can we conclude from the 95% CI in the problem above?

a. There is no statistically significant difference in the percentage with acne among those age 21-30 and those age 31-40.

b. There is statistically significant correlation between the sample size and the effect size for the confounding factors of a randomized controlled experiment.

c. Of those with acne, there is no statistically significant difference between the percentage who are age 21-30 and the percentage who are 31-40.

d. A statistically significantly higher percentage of people age 21-30 have acne than those age 31-40.

e. The sample sizes are too small for a conclusion to be valid here.

Suppose a researcher studying acne, or oily skin, takes a simple random sample of 400 Americans age 21-30 and calls them group A, and a simple random sample of 200 Americans age 31-40 and calls them group B. In group A, 88 people have acne, and in group B, 30 people have acne.

13. What can we conclude from the 95% CI in the problem above?

a. There is no statistically significant difference in the percentage with acne among those age 21-30 and those age 31-40.

b. There is statistically significant correlation between the sample size and the effect size for the confounding factors of a randomized controlled experiment.

c. Of those with acne, there is no statistically significant difference between the percentage who are age 21-30 and the percentage who are 31-40.

d. A statistically significantly higher percentage of people age 21-30 have acne than those age 31-40.

e. The sample sizes are too small for a conclusion to be valid here.

For the following two problems, suppose a researcher studies a treatment for a certain condition. The researcher divides 230 subjects randomly into three groups. Group A receives a placebo. Group B receives the treatment at a low dose, and group C receives the treatment at a high dose. The table below shows the output from an ANOVA F-test on the mean levels of the condition in the 3 groups.

Analysis of Variance Table

Response: outcome

	Df	Sum Sq	MeanSq	F value	Pr(>F)
group	2	0.467	0.233	0.221	0.803
Residuals	227	28.5	1.06		

14. In the table, what number is a measure of the variability between groups?

a. 2. b. 0.233 c. 0.221 d. 0.803 e. 28.5 f. 1.06.

For the following two problems, suppose a researcher studies a treatment for a certain condition. The researcher divides 230 subjects randomly into three groups. Group A receives a placebo. Group B receives the treatment at a low dose, and group C receives the treatment at a high dose. The table below shows the output from an ANOVA F-test on the mean levels of the condition in the 3 groups.

Analysis of Variance Table

Response: outcome

	Df	Sum Sq	MeanSq	F value	Pr(>F)
group	2	0.467	0.233	0.221	0.803
Residuals	227	28.5	1.06		

14. In the table, what number is a measure of the variability between groups?

a. 2. **b. 0.233** c. 0.221 d. 0.803 e. 28.5 f. 1.06.

For the following two problems, suppose a researcher studies a treatment for a certain condition. The researcher divides 230 subjects randomly into three groups. Group A receives a placebo. Group B receives the treatment at a low dose, and group C receives the treatment at a high dose. The table below shows the output from an ANOVA F-test on the mean levels of the condition in the 3 groups.

Analysis of Variance Table

Response: outcome

	Df	Sum Sq	MeanSq	F value	Pr(>F)
group	2	0.467	0.233	0.221	0.803
Residuals	227	28.5	1.06		

15. What can we conclude from the result of the F test?

a. The treatment has a statistically significant effect, though we cannot be sure which dose is attributable to this significant effect or in which direction the effect goes.

b. The treatment seems to have significantly greater effect at large doses than at small doses.

c. The treatment does not seem to have a statistically significant effect.

d. We fail to reject the null hypothesis that the treatment has a significant effect on the condition.

e. The correlation between dose and outcome appears to be statistically significant.

For the following two problems, suppose a researcher studies a treatment for a certain condition. The researcher divides 230 subjects randomly into three groups. Group A receives a placebo. Group B receives the treatment at a low dose, and group C receives the treatment at a high dose. The table below shows the output from an ANOVA F-test on the mean levels of the condition in the 3 groups.

Analysis of Variance Table

Response: outcome

	Df	Sum Sq	MeanSq	F value	Pr(>F)
group	2	0.467	0.233	0.221	0.803
Residuals	227	28.5	1.06		

15. What can we conclude from the result of the F test?

a. The treatment has a statistically significant effect, though we cannot be sure which dose is attributable to this significant effect or in which direction the effect goes.

b. The treatment seems to have significantly greater effect at large doses than at small doses.

c. The treatment does not seem to have a statistically significant effect.

d. We fail to reject the null hypothesis that the treatment has a significant effect on the condition.

e. The correlation between dose and outcome appears to be statistically significant.

______ 16. For the next two problems, a researcher studies the effect of watching a scary movie on heart rate. She takes a simple random sample of 200 subjects, and randomly divides them into two groups. Group A is shown 10 minutes of a scary movie, and their heart rates are recorded. Group B is shown 10 minutes of a non-scary movie, and their heart rates are recorded. The next week, the subjects return, and Group A is shown 10 minutes of a non-scary movie, Group B is shown 10 minutes of a scary movie and for both groups their heart rates are recorded. The researcher finds the following, for the 200 subjects in groups A and B combined:

Mean heart rate when watching scary movie: 88 bpm.
Mean heart rate when watching non-scary movie: 86 bpm.
SD of heart rates when watching scary movie: 22 bpm.
SD of heart rates when watching non-scary movie: 20 bpm.
SD of differences (scary movie heartrate - non-scary movie heartrate): 8 bpm.

Which of the following is an appropriate Z-statistic for testing whether or not the observed difference in heart rates is statistically significant?
a. 0.572. b. 0.998. c. 1.31. d. 1.77. e. 3.54. f. None of the above.

______ 16. For the next two problems, a researcher studies the effect of watching a scary movie on heart rate. She takes a simple random sample of 200 subjects, and randomly divides them into two groups. Group A is shown 10 minutes of a scary movie, and their heart rates are recorded. Group B is shown 10 minutes of a non-scary movie, and their heart rates are recorded. The next week, the subjects return, and Group A is shown 10 minutes of a non-scary movie, Group B is shown 10 minutes of a scary movie and for both groups their heart rates are recorded. The researcher finds the following, for the 200 subjects in groups A and B combined:

Mean heart rate when watching scary movie: 88 bpm.
Mean heart rate when watching non-scary movie: 86 bpm.
SD of heart rates when watching scary movie: 22 bpm.
SD of heart rates when watching non-scary movie: 20 bpm.
SD of differences (scary movie heartrate - non-scary movie heartrate): 8 bpm.

Which of the following is an appropriate Z-statistic for testing whether or not the observed difference in heart rates is statistically significant?
a. 0.572. b. 0.998. c. 1.31. d. 1.77. **e. 3.54.** f. None of the above.

(88-86)/(8/sqrt(200)) ~ 3.54.

______ 17. Continuing the previous problem, suppose you wanted to use simulations to find a null distribution for your standardized statistic. Which of the following would be an appropriate way to conduct the simulations?

a. For each subject, flip a coin, and if the coin is heads, interchange the subject's scary movie heartrate with the subject's non-scary movie heartrate, and otherwise keep the subject's scores the same.

b. For each subject, flip a coin, and if the coin is heads, interchange the subject's scary movie heartrate with another subject's scary movie heartrate, where the other subject is selected at random among the 199 other subjects.

c. For each subject, flip a coin, and if the coin is heads, interchange the subject's scary movie heartrate with the sample mean heartrate for all 200 subjects.

d. For each subject, flip a coin, and if the coin is heads, interchange the subject's non-scary movie heartrate with 86 bpm.

e. For each subject, flip a coin, and if the coin is heads, interchange the subject's non-scary movie heartrate with 86 bpm +/- 20 bpm.

______ 17. Continuing the previous problem, suppose you wanted to use simulations to find a null distribution for your standardized statistic. Which of the following would be an appropriate way to conduct the simulations?

a. For each subject, flip a coin, and if the coin is heads, interchange the subject's scary movie heartrate with the subject's non-scary movie heartrate, and otherwise keep the subject's scores the same.

b. For each subject, flip a coin, and if the coin is heads, interchange the subject's scary movie heartrate with another subject's scary movie heartrate, where the other subject is selected at random among the 199 other subjects.

c. For each subject, flip a coin, and if the coin is heads, interchange the subject's scary movie heartrate with the sample mean heartrate for all 200 subjects.

d. For each subject, flip a coin, and if the coin is heads, interchange the subject's non-scary movie heartrate with 86 bpm.

e. For each subject, flip a coin, and if the coin is heads, interchange the subject's non-scary movie heartrate with 86 bpm +/- 20 bpm.

Suppose a researcher takes a simple random sample of 100 high school students, and finds that the number of books read per year per student has a mean of 7.2, a median of 7.0, a 25th percentile of 2.0, and a 75th percentile of 10.0. She finds a 95% CI for the mean number of books read per year per student among high school students in general is 7.2 +/- 1.0.

______ 18. Find s, the sample standard deviation of the number of books read per year among the 100 students in the sample.

a. 5.10.   b. 7.23.   c. 7.90.   d. 8.24.   e. 8.81.   f. None of the above.

Suppose a researcher takes a simple random sample of 100 high school students, and finds that the number of books read per year per student has a mean of 7.2, a median of 7.0, a 25th percentile of 2.0, and a 75th percentile of 10.0. She finds a 95% CI for the mean number of books read per year per student among high school students in general is 7.2 +/- 1.0.

______ 18. Find s, the sample standard deviation of the number of books read per year among the 100 students in the sample.

a. 5.10. b. 7.23. c. 7.90. d. 8.24. e. 8.81. f. None of the above.

$1.0 = 1.96 * s / \sqrt{100}$, so $s = \sqrt{100} / 1.96 = 5.10$.