BIO 211

Test Pac

$$s^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1} = \frac{SS}{DF}$$

The sample variance

Version 12.0

Revision Date: July 2009

David J. Moriarty
Biological Sciences Department
California State Polytechnic University

The BIO 211 TEST PAC contains a variety of information critical to your success in the course. Bring the TEST PAC to class with you, and be sure to bring it to exams. Use the TEST PAC regularly and become very familiar with its contents. It will help you.

Information in the TEST PAC includes: (1) test summary sheets; (2) class examples; (3) supplemental reading material; (4) practice exams; and (5) practice problems. These are discussed below. See the Table of Contents to locate various items.

Test Summary Sheets

The test summary sheets contain a summary of basic information on all the inferential statistical tests covered in the course. **You are required to have this information committed to memory**. Much of this information will appear on the closed book portion of exams.

Another important use for this information is to help you select the appropriate statistical test on the open book portion of the exam. When you read a problem, do not expect the choice of test needed to solve the problem to be obvious. You must have a systematic method of going through the tests available to you, and selecting the correct one. Remember that the choice of tests depends on two factors: 1) the biological question; and 2) characteristics of the data. The test summaries should aid you in your search for the appropriate test. You are allowed (and expected) to bring the TEST PAC to exams and use it on the open book portion of the exams.

While these sheets present a summary of many important facts on the various analyses, they are not complete. Do not attempt to substitute these sheets for class attendance.

Class Examples

The examples used in class for many of the statistical tests are found in the TEST PAC. We will use the examples extensively in class. You'll need the TEST PAC!

Supplemental Reading Material

The TEST PAC contains text on critical areas are not adequately covered in the text. Your course syllabus references these reading assignments.

Practice Exams

Read the discussion on how to use the practice exams about one week prior to the first exam. See the Table of Contents for the section of practice exams.

Practice Problems - Werner Blood Chemistry Data

The practice problems are designed to help you on the final exam. Read the discussion on the practice problems a few days before the final.

Basic Statistics Review

Students are strongly encouraged to take a course in basic statistics (STA 120 or equivalent) prior to taking Biometrics. A basic statistics course not only covers fundamental material critical for success in Biometrics, but also provides an extensive introduction to topics such as probability, which are important in the biological sciences. You will find review material for the parts of basic statistics critical to BIO 211 on the course Blackboard site.

Table of Contents

A GENERAL METHOD OF HYPOTHESIS TESTING

How do inferential statistical tests work? How are hypotheses tested? An understanding of the answers to these questions is critical for success in biometrics. This topic will be the subject of a lecture in class (see your lecture schedule under Hypothesis Testing). This section of the Test Pac is designed to give you some notes and reading for the lecture. This topic tends to be confusing and somewhat difficult. Try to follow the lecture closely; read and refer to these notes; and ask questions!

All of the inferential tests covered in this class share a general methodology. The details of how that methodology is implemented are often quite different, but the general approach is the same.

When analyzing data, we want to use that data to answer a biological question. We want to use our sample estimates to make some inference about the biological populations involved. For example, we may have estimates of the average height of students at Cal Poly, and the average height of students at Cal State LA. We want to use these estimates to determine if there is a difference between Cal State LA and Cal Poly in the height of students.

The method used is this:

1. Assume populations are the same with respect to the parameter, i.e. assume no difference. The statement of this no difference hypothesis (assumption) is called the **null hypothesis**.

2. When we examine our sample estimates (descriptive statistics) we see they are different. But does this mean that the populations are different? The differences seen in the samples could be explained in two ways:

   A. The populations are different, and that's why their estimates are different. Our initial assumption in number 1 above (our null hypothesis) is wrong and should be rejected.

   B. The populations are the same, and the difference seen in the samples is just due to random sampling error. Our initial assumption (null hypothesis) is correct and should be accepted.

   We must now decide which explanation (A or B) is correct.

3. We now ask how much difference there is in our samples. We want to quantify the difference. **Inferential statistics are numbers that quantify differences.**

4. Is the difference a big difference or a small difference? A small difference could happen just by random sampling error, so if the difference is small we will use explanation B above. A big difference is unlikely to occur just by chance, so if the difference is big, our explanation will be choice A.

5. To determine if we have a big or small difference, we ask what is the probability of obtaining this much difference just by chance if we have sampled no difference populations (i.e. if our null hypothesis in 1 is correct). This probability is called "alpha probability".

   A small difference has a large probability (>0.05) of occurring. A big difference has a small probability (≤ 0.05) of occurring.

   Since the inferential statistic quantifies the difference, we must determine the probability of finding the particular value of the

statistic. The sampling distribution of the statistic allows us to determine the probability.

If the alpha probability of the statistic is >0.05, we use explanation B. That is, the null hypothesis is accepted.

If the alpha probability of the statistic is ≤0.05, we use explanation A. That is, the null hypothesis is rejected.

When we accept or reject a null hypothesis, we hope we are making the right decision. However, there is always some probability of us being wrong. Notice that there are two possible ways in which we might be wrong:

1. We might reject a null hypothesis that we should have accepted, that is, we conclude there is a difference present when there really isn't. Statisticians call this a **Type I** error. When we reject a null hypothesis, the probability that we have committed a Type I error is the alpha probability (just as defined above).

2. We might accept a null hypothesis that we should have rejected. In this case, we have failed to find a difference that actually does exist. Statisticians call this a **Type II** error. When we accept a null hypothesis, the probability that we have committed a Type II error is called beta probability. Beta probability is difficult to calculate, because the probability of missing some difference depends on how big the difference is, which we can never know for certain. The ability of a statistical test to avoid making a Type II error is called the **power** of the test. In other words, **power** refers to how well the test can detect a difference. A powerful test is one that can detect small differences.

The table below summarizes Type I and Type II errors:

	You should have Rejected Ho:	You should have Accepted (not Rejected) Ho:
You Rejected Ho:	☺ You are correct!	☹ Type I error (α probability)
You Accepted (failed to Reject) Ho:	☹ Type II Error (β probability)	☺ You are correct!

With respect to alpha probability (α) and beta probability (β), it is important to realize that $\beta \neq 1 - \alpha$. As discussed above, we get α probabilities from the tables in the back of our book. However, we don't usually know β, because it would require us to know how different the populations really are - and we never know that!

When we make scientific conclusions, we want to be correct. In other words, we want to have both α and β be the smallest values possible. α and β are inversely related, i.e. as one goes up the other goes down. Statisticians have shown both theoretically and empirically, that (as a general rule) you can minimize both α and β by using an α value of about 0.05. If you use a smaller α, the β goes up too high. This is why statisticians generally recommend that null hypotheses be rejected at the α=0.05 value. Although it may seem like an arbitrary value to us biologists, there are actually good mathematical reasons for using 0.05.

The only way to simultaneously decrease both α and β is to increase your sample size.

GOODNESS-OF-FIT

WHAT IS TESTED
A difference between a set of observed frequencies and a set of expected frequencies, where the expected frequencies come from an *a priori* ratio or distribution. The phrase *a priori* means from some source other than the observed data.

DATA SCALE
Nominal

NULL HYPOTHESIS
Ho: No difference between the observed frequencies and those expected from the *a priori* ratio or distribution.

ASSUMPTIONS WITH RESPECT TO DATA
None

DEGREES OF FREEDOM
DF = k - 1 where k is the number of categories (i.e. the number of expected frequencies). This formula (DF = k - 1) is correct for all the applications used in this class. You should be aware that the formula may be different in other applications.

ZAR CHAPTER
22

TYPE
Nonparametric

COMMENTS
We will observe "Roscoe and Byars Rule": The average *expected* frequency must not be below 2.0. You can easily and quickly calculate the average expected frequency by dividing the sample size (n) by the number of categories (k).

Also beware of the Yates' correction for continuity, which should be used when there is 1 degree of freedom (see Zar).

Never use percentages or proportions as data for this test (or for any other test using nominal scale data). Convert percentages or proportions to actual frequencies, and then perform the test on the frequencies.

This is not a powerful test (which means that it is not very good at detecting small differences). However, there is no other procedure available which is applicable in as many situations as the chi-squared approach.

Heterogeneity testing (discussed in Zar) is used to determine if several samples come from the same population. If the answer is yes (null accepted), then samples may be pooled to form one large sample. You need to be aware of what heterogeneity testing accomplishes, you don't have to know how to do one (although it's pretty easy).

GOODNESS-OF-FIT TEST - Example

In the garden pea plant (*Pisum sativum*), yellow seed color (Y) is dominant to green (y); and round seed (R) is dominant to wrinkled (r). In a dihybrid cross (i.e. parents are heterozygous for both traits), the offspring would be expected to have a 9:3:3:1 phenotypic ratio:

9/16 Yellow, Round (Y_R_)
3/16 Yellow, wrinkled (Y_rr)
3/16 green, Round (yyR_)
1/16 green, wrinkled (yyrr)

Gregor Mendel (1822-1884) reported the results of a dihybrid cross which resulted in 556 offspring. He found 315 Yellow, Round; 101 Yellow, wrinkled; 108 green, Round; and 32 green, wrinkled. Are his results consistent with the theoretical expectation from a dihybrid cross?

Note that expected frequencies are calculated by multiplying the expected proportion times the total sample size. In this example:

Yellow, Round: 9/16 x 556 = 312.75
Yellow, wrinkled: 3/16 x 556 = 104.25
green, Round: 3/16 x 556 = 104.25
green, wrinkled: 1/16 x 556 = 34.75

Ho: No difference between the observed phenotypic frequencies and those expected from a 9:3:3:1 ratio.

	f_i	$\hat{f}_i$	$\frac{\left(f_i - \hat{f}_i\right)^2}{\hat{f}_i}$
Yellow, Round	315	312.75	0.016
Yellow, wrinkled	101	104.25	0.101
green, Round	108	104.25	0.135
green, wrinkled	32	34.75	0.218
	---	------	-----
	n = 556	556	χ^2 = 0.470

DF = k - 1 where k is the number of categories. k=4 in this example.
DF = 4 - 1 = 3

Using Table B.1 (Zar 4th edition page App12; Zar 5th edition page 672), we find that p > 0.05, therefore Accept Ho: (fail to reject Ho:).

These results are consistent with a dihybrid cross.

Notice that the observed and expected frequencies *must* sum to the same value (556 in this example).

Check for biases or corrections:

1. Continuity correction.

   This is applicable **only when DF=1**. Since we have DF=3 in our example above, no continuity correction is needed.

   When you have a problem with only two categories (k=2), then you have DF = 1. Use the Yates' Correction for Continuity, which is this formula:

$$\chi_c^2 = \sum_{i=1}^{2} \frac{(\left|f_i - \hat{f}_i\right| - 0.5)^2}{\hat{f}_i}$$

   Be sure to practice this using the example in your textbook.

   Remember, the correction is only applied when DF = 1.

2. Sample size.

   This test is biased if the sample size is too low. In order to determine if the sample size is too low, we will apply a simplified version of what might be called "Roscoe and Byars Rule": The average *expected* frequency must not be below 2.0

   You can easily and quickly calculate the average expected frequency by dividing the sample size (n) by the number of categories (k).

   In our example, our n = 556, and k = 4. Therefore, the average expected frequency is 556/4 = 139, which is a lot greater than 2.0. Therefore, we have no sample size problem in our example.

   Notice that a sample size of n=8 would be the minimum required (8/4 = 2). See Zar for a complete discussion of the sample size bias issue.

CONTINGENCY TABLE ANALYSIS

WHAT IS TESTED
Testing to see if the frequencies of occurrence of the categories of one variable are independent of the frequencies of the categories of a second variable. Expected frequencies are *a posteriori*, i.e. they come from the observed data.

DATA SCALE
Nominal

NULL HYPOTHESIS
Ho: Rows and columns are independent.

ASSUMPTIONS WITH RESPECT TO DATA
None

DEGREES OF FREEDOM
DF = (r-1)(c-1), where r is the number of rows in the table and c is the number of columns.

ZAR CHAPTER
23

TYPE
Nonparametric

COMMENTS
Watch for "Roscoe and Byars Rule", which states that the average expected frequency should be at least 6.0. Calculate the average expected frequency as n/(r×c).

When there is 1 degree of freedom (i.e. a 2X2 table), a correction for continuity must be applied. You may use the Yates' correction for continuity (Zar 4th edition pages 493-494, example 23.2; Zar 5th edition pages 500-502, example 23.4), which is applied in a similar fashion as in the Goodness-of-fit test. You may also use the Haber correction (Zar 4th edition pages 494-495; example 23.3). In Zar 5th edition, it is called the Cochran-Haber correction (pages 501-502, example 23.4). The Haber (Cochran-Haber) correction is actually simple, although confusing the first time you try to follow the example.

Never use percentages or proportions as data for this test (or for any other test using nominal scale data). Convert percentages or proportions to actual frequencies, and then perform the test on the frequencies.

Heterogeneity testing (Zar 4th edition, pages 500-502; Zar 5th edition pages 504-506) is used to see if several samples come from the same population. If the answer is yes (null accepted), then samples may be pooled to form one large sample. Again, you won't have to do a heterogeneity test.

We will NOT distinguish among Category 1, Category 2, and Category 3 tables as discussed in Zar 4th edition on pages 491-499. In Zar 5th edition, pages 497-500, these are called (a) No Margin Fixed, (b) One Margin Fixed, (c) Both Margins Fixed.

For small sample sizes in contingency tables, see the Fisher Exact Test (Zar, Chapter 24 – not in your assigned reading). The Fisher Exact Test may be used even when the sample size is very low (i.e. so low that it violates the "Roscoe and Byars Rule"). In this test, you don't calculate chi-squared, but rather you use exact binomial/multinomial probabilities. Although Zar only discusses the Fisher Exact Test in the context of 2x2 tables, it can be done on higher dimension tables. You won't be asked to do the Fisher Exact Test in this course, but you need to be aware of it. The Fisher Exact Test is commonly seen in the biological literature.

CONTINGENCY TABLE ANALYSIS - Example

Over a 2-year period, 327 patients with hypertension were studied at a hospital. Patients were randomly assigned to one of three drug types (A, B, or C), and the drugs administered for 3 months. At the end of the 3 month period, a physician categorized each patient as "No Change" or "Improved". Note: during the course of the study, 27 patients were determined to be getting worse. They were withdrawn from the study and treated by other means. We will analyze only the 300 patients who completed the study. Withdrawing patients from a clinical study for medical reasons is a common and necessary part of many research protocols.

	No Change	**Improved**	**Row Totals**
Drug A	20	20	40
Drug B	40	60	100
Drug C	40	120	160
Column Totals	100	200	300 = n

Ho: No difference between the observed frequencies and those expected if response (the "No Change" to "Improved" ratio) is independent of drug type.

Expected frequencies in row i and column j are calculated by multiplying the row total for row i (R_i) times the column total for column j (C_j,) and then dividing by n. Below is the formula and the calculation shown for the Drug A, No Change cell.

$$\hat{f}_{ij} = \frac{R_i C_j}{n} \qquad \hat{f}_{11} = \frac{40 \times 100}{300} = 13.33$$

In the table below, the expected frequencies for each cell are in parentheses.

	No Change	**Improved**	**Row Totals**
Drug A	20 (13.33)	20 (26.67)	40
Drug B	40 (33.33)	60 (66.67)	100
Drug C	40 (53.33)	120 (106.67)	160
Column Totals	100	200	300 = n

The chi-squared statistic is calculated by taking (observed-expected)^2/expected for each cell, and then summing across all six cells. The formula is

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{\left(f_{ij} - \hat{f}_{ij} \right)^2}{\hat{f}_{ij}} \quad r = number\ of\ rows; \quad c = number\ of\ columns$$

The double summation sign is necessary because of the subscripts for the rows and columns.

The actual calculations are below:

Cell	Observed	Expected	$(\text{Observed} - \text{Expected})^2 / \text{Expected}$
Drug A, No Change	20	13.33	3.34
Drug A, Improved	20	26.67	1.67
Drug B, No Change	40	33.33	1.33
Drug B, Improved	60	66.67	0.67
Drug C, No Change	40	53.33	3.33
Drug C, Improved	120	106.67	1.67
			$\chi^2 = 12.01$

In this example, you should get $\chi^2 = 12.01$.

DF = (r-1)(c-1) where r is the number of rows and c is the number of columns.

In our example, DF = (3-1)(2-1) = 2.

Consulting the chi-squared table (Zar 4th ed. pg. App12; 5th ed. pg. 672) we find that $p < 0.05$, therefore we reject Ho:.

The proportion of patients showing improvement depends on drug type.

Notice that a correction for continuity is not needed because the DF is not one.

Our average expected frequency is 50 (300/(3×2) = 50), so we have sufficient sample size. (The "Roscoe and Byars Rule" for contingency tables is that the average expected frequency must be at least 6.0).

ONE-SAMPLE *t*-TEST

WHAT IS TESTED
A difference between the mean of a single sample and some constant value (not the mean of another sample) which is generated *a priori*. The test determines the probability that the sample is a random sample from a population with a mean equal to the constant value.

DATA SCALE
Ratio-Interval

NULL HYPOTHESIS
Ho: $\mu = c$ where c is the constant value

One-tailed hypotheses may be tested.

ASSUMPTIONS WITH RESPECT TO DATA
The data sample is a RSNDP (Random Sample from a Normally Distributed Population). The RSNDP abbreviation will appear throughout this Test Pac.

DEGREES OF FREEDOM
DF = n - 1 where n is the sample size.

ZAR CHAPTER
7

TYPE
Parametric. There is no nonparametric analogue.

COMMENTS
This test is robust, i.e. it can withstand moderate deviations from the assumption without any important adverse effect, particularly with large sample sizes (n>25).

One-tailed and two-tailed null hypotheses can be tested.

Read about (and know how to calculate) confidence limits for the mean:
Zar 4th edition: pages 98 – 100
Zar 5th edition: pages 105 - 107

One-sample *t*-test - Example

While browsing through a book on veterinary medicine, we encounter the statement "Horses live an average of 22 years". There are no data given, just the single number of 22. In other words, 22 is a *constant*. We wonder if this statement is correct, so we gather longevity data on 25 horses. We test to see if these 25 horses are a random sample from a population whose mean is 22.

Ho: μ = 22 yrs

	Age at Death (yrs)
Horse 1	17.2
Horse 2	18.0
Horse 3	18.7
Horse 4	19.8
Horse 5	20.3
Horse 6	20.9
Horse 7	21.0
Horse 8	21.7
Horse 9	22.3
Horse 10	22.6
Horse 11	23.1
Horse 12	23.4
Horse 13	23.8
Horse 14	24.2
Horse 15	24.6
Horse 16	25.8
Horse 17	26.0
Horse 18	26.3
Horse 19	27.2
Horse 20	27.6
Horse 21	28.1
Horse 22	28.6
Horse 23	29.3
Horse 24	30.1
Horse 25	35.1
Mean	24.23 yrs
Standard Deviation (s)	4.25 yrs
n	25

Note: If you do not know how to calculate the standard deviation (s), it's time to review your basic statistics! See: Basic Statistics Part One, available on Blackboard or at the class web site.

$$t = \frac{\overline{X} - c}{s_{\overline{X}}} \text{ where } c \text{ is the constant}, \quad \text{and } s_{\overline{X}} = \frac{s}{\sqrt{n}}$$

$$\text{therefore, } s_{\overline{X}} = \frac{4.25}{\sqrt{25}} = 0.85 \quad \text{and} \quad t = \frac{24.23 - 22}{0.85} = 2.624$$

DF = n - 1 = 25 - 1 = 24. Consulting the *t* table (Zar 4th edition page App19; Zar 5th edition page 678) we find that a value of 2.624 with 24 DF has a two-tailed alpha probability $p<0.05$. Therefore, we reject the null hypothesis. We conclude that our sample of horses does not come from a population that averages 22 years. Perhaps the book is incorrect, or maybe our sample is biased for some reason. It's time to do some biology!

ONE-TAILED VS. TWO TAILED TESTS

Many statistical tests can be done using either one-tailed or two-tailed null hypotheses.

The following statistical tests covered in this course can be done either one- or two-tailed: One-sample *t*-test; Variance Ratio Test; Two-sample *t*-test; Mann-Whitney U Test; Paired-sample *t*-test; Wilcoxon Paired Sample Test; Regression; Correlation; Spearman Rank Correlation.

The concept of one-tailed vs. two-tailed hypotheses will not be applied to the following tests: Goodness-of-Fit; Contingency Table Analysis; Bartlett's Test; ANOVA (all models); Multiple Comparisons; Kruskal-Wallis Test; ANCOVA.

As the concept of one-tailed vs. two-tailed hypotheses is discussed below, the One-sample *t*-test will be used as an example. However, you need to be aware that the concept is applied to other tests we will learn as indicated above.

One-tailed null hypotheses are used when there is a biological expectation of a difference *in a particular direction*. For example, suppose I read in the veterinary medical literature that horses live an average of 22 years. If, based on my biological knowledge, I believe that is too low, a one-tailed hypothesis would be appropriate. My biological expectation is that the actual value of the population mean is greater than 22. *One-tailed null hypotheses must always be justified biologically. If you don't have an expectation of a difference in a particular direction, or are unsure, use the two-tailed null hypothesis. The biological expectation must be* a priori*, that is, you must have the expectation before you see the data. It is wrong to test a one-tailed null hypothesis based on what you see in the data.*

There are always two possible one-tailed null hypotheses. Which one you test depends on the direction you expect to find in the difference. For our horse problem in the One-sample *t*-test, the two possible one-tailed null hypotheses are:

$H_o: \mu \leq 22$
$H_o: \mu \geq 22$

The two-tailed null hypothesis is the one you are already know, because we tested this hypothesis in class. The two-tailed null is $H_o: \mu = 22$.

Which one-tailed null hypothesis you test depends on your expectation. BE CAREFUL! This can be tricky! Let's use the horse longevity example.

If your expectation is that horses actually live longer than 22 years, the correct null hypothesis is $H_o: \mu \leq 22$.

If your expectation is that horses actually live less than 22 years, the correct null hypothesis is $H_o: \mu \geq 22$.

Do you understand why? The answer involves how we "prove" a hypothesis. In the scientific method, we don't directly prove a hypothesis to be true. What we do is state all of the possible hypotheses, and begin testing them. When we have rejected all but one, the one remaining is accepted as "truth". Let's apply this to the horse problem.

Suppose I expect that horses actually live longer than 22 years. To "prove" this, what I have to do is state all possible hypotheses about the population mean of horse longevity and the constant of 22. All possible hypotheses are:

$\mu < 22$
$\mu = 22$
$\mu > 22$

Since I expect that horses live longer than 22 years ($\mu > 22$), I must reject the other possibilities. Therefore, my null hypothesis is Ho: $\mu \leq 22$.
If your expectation was that horses live less than 22 years, you would have to test the null hypothesis Ho: $\mu \exists 22$. If you rejected this null, you would have "proven" that $\mu < 22$, i.e. horses live less than 22 years.

Another way to approach this is to remember that your expected difference is always the *alternate* hypothesis (H_A:). In your textbook, Zar always states both the null and the alternate hypothesis.

For example, if you expect horses live longer than 22 years, then:
Ho: $\mu \leq 22$
H_A: $\mu > 22$

If you expect horses live less than 22 years, then:
Ho: $\mu \geq 22$
H_A: $\mu < 22$

Be careful! Thinking is required!

Once you have stated a one-tailed null hypothesis, use the $\alpha(1)$ probabilities in the statistical tables in Zar. Always use $\alpha(1)$ probabilities when testing a one-tailed null hypothesis. Always use $\alpha(2)$ probabilities when testing a two-tailed null hypothesis.

Another place where you need to be careful is to make sure your difference is in the appropriate direction. In a one-tailed test, the difference must be in the direction that matches your expectation, or you must accept the null.

For example, when testing the Ho: $\mu \leq 22$, we must find a sample mean GREATER than 22 if we are going to reject. If our sample mean turns out to be less than or equal to 22, we accept immediately, we don't even calculate *t*.

Since the formula for *t* is: $$t = \frac{\overline{X} - c}{s_{\overline{X}}}$$

the sign (+ or -) on *t* is important. If the null hypothesis is Ho: $\mu \leq 22$, *t* would have to be positive (and $p<0.05$) for us to reject. Note that *t* will be positive when the sample mean is greater than the constant.

If the null hypothesis is Ho: $\mu \geq 22$, *t* would have to be negative and ($p<0.05$) for us to reject. *t* will be negative when the sample mean is less than the constant.

Again, BE CAREFUL! THINK!

In summary:

1. Recognize one-tailed biological questions. They always ask about differences in a particular direction. Examine the problem carefully for concepts such as greater, less, more, above, below, positive, negative. Do *not* try to use a "key word" approach (i.e. if certain words are present, then I'll do a one-tailed test). You can *not* do this with key words - unfortunately you must read the problem carefully and *think* about what you're doing - just like an actual scientist!

2. If a one-tailed null hypothesis is appropriate, be sure you state the correct one!

3. Make sure the difference is in the appropriate direction before you reject the null hypothesis!

VARIANCE RATIO TEST

WHAT IS TESTED
A difference between the variances of two populations.

DATA SCALE
Ratio-Interval

NULL HYPOTHESIS
Ho: $\sigma_1^2 = \sigma_2^2$

One-tailed hypotheses may be tested.

ASSUMPTIONS WITH RESPECT TO DATA
Both samples are a RSNDP.

DEGREES OF FREEDOM
Two degree of freedom values are necessary, one for the numerator, and one for the denominator of the F value. In each case, the degree of freedom is calculated as n-1, where n is the appropriate sample size. The appropriate sample size is the sample size corresponding to the variance. For example, if the variance from sample 1 is used in the numerator, then the degree of freedom for the numerator is the size of sample 1 minus 1. The degree of freedom for the denominator is the size of sample 2 minus 1.

ZAR CHAPTER
8

TYPE
Parametric. There is no nonparametric analogue.

COMMENTS
This test is not robust, but is severely and adversely affected by samples that are not normally distributed. A procedure called Levene's Test may be used as an alternative to the Variance Ratio Test. Levene's Test also assumes RSNDP, but is more robust to deviations from the normality assumption.

Another approach to dealing with samples that are not normally distributed is to use resampling procedures, e.g. the jackknife, or the bootstrap, or randomization methods. Resampling procedures are computationally intensive, and should be done on computers.

When calculating the F ratio for a two-tailed test, divide the larger sample variance by the smaller. For one-tailed null hypotheses, divide the sample variance you *expect* to be larger by the one you *expect* to be smaller.

In terms of formulas:

For Ho: $\sigma_1^2 = \sigma_2^2$ $\quad F = \dfrac{\text{larger } s_i^2}{\text{smaller } s_i^2}$

For Ho: $\sigma_1^2 \leq \sigma_2^2$ $\quad F = \dfrac{s_1^2}{s_2^2}$ $\qquad$ For Ho: $\sigma_1^2 \geq \sigma_2^2$ $\quad F = \dfrac{s_2^2}{s_1^2}$

TWO-SAMPLE *t*-TEST

WHAT IS TESTED
A difference between the means of two populations.

DATA SCALE
Ratio-Interval

NULL HYPOTHESIS
Ho: $\mu_1 = \mu_2$

One-tailed hypotheses may be tested.

ASSUMPTIONS WITH RESPECT TO DATA
Both samples are RSNDP. Samples are from populations with equal variances (homoscedasticity).

DEGREES OF FREEDOM
$DF = n_1 + n_2 - 2$

ZAR CHAPTER
8

TYPE
Parametric. Nonparametric analogue is the Mann-Whitney U.

COMMENTS
This is a robust analysis.

When data meet the assumption of normality, but not of equal variances, there is an alternate parametric procedure.

In Zar, 4th edition, this procedure is called "Welch's approximate *t*", and is discussed on pages 128-129.

In Zar, 5th edition, this procedure is called "Welch's approximate *t*", or the "Behrens-Fisher Test", and is discussed on pages 137-141.

Do not worry about how to do this procedure, just know that it exists.

The assumption of equal variances can be tested with the Variance Ratio Test. In this course, we will test the homoscedasticity assumption with the variance ratio before doing the Two-sample *t*-test. If the null hypothesis of equal variances is rejected in the Variance Ratio Test, then use the Mann-Whitney U rather than the Two-sample *t*-test.

Variance Ratio & Two-sample t-test - Example

A patient is scheduled to begin taking a new medication in February. In order to assess the effect of the medication on blood glucose level, the patient's blood glucose is measured at 8:00 am for 14 days in January and 13 days in February. Blood glucose is measured in mg/dl.

January	February
74	97
97	94
95	92
93	89
85	90
85	95
91	92
92	94
95	94
88	93
97	99
92	96
95	98
96	
Mean 91.1	94.1
Standard Deviation (s) 6.33	2.96
Variance (s^2) 40.07	8.76
n 14	13

Variance Ratio Test

We are concerned that the medication might affect the variability in blood glucose levels. We do not have an *a priori* expectation of how the variability might be affected, i.e. we don't know if it will go up, go down, or stay the same.

Ho: $\sigma_1^2 = \sigma_2^2$ $$F = \frac{s_1^2}{s_2^2} = \frac{40.07}{8.76} = 4.57$$

$DF_{numerator} = n_1 - 1 = 14 - 1 = 13$ $DF_{denominator} = n_2 - 1 = 13 - 1 = 12$

Consulting the F table in Zar for our degrees of freedom (4th edition: page App33; 5th edition: page 692), we find that $p<0.05$, therefore we reject Ho:. We have found that blood glucose is more variable in January than in February.

Do you think this is a statistically and biologically correct conclusion?

Examine the data carefully, and think about the assumption (RSNDP) of the Variance Ratio Test. Do you see any potential problems?

Read the next page!

Variance Ratio & Two-sample t-test - Example (continued)

Note that the first datum for January on the previous page is 74. Visual inspection of the other data indicates this value may be a negative outlier, i.e. it is much smaller than the other values. This may be affecting the variance ratio test, perhaps by making the data nonnormal. The variance test is known to perform poorly (i.e. give you the wrong conclusion) when there are outliers in the data. This data set would be a candidate for the Levene's test or a resampling procedure to test the variances.

Another way to investigate our suspicion that the single data point is affecting our conclusion is to do the analysis without the datum. If you remove the value of 74 from the January data, the mean is 92.384, the variance is 17.256, and n=13. Notice that the variance has dropped from over 40 to about 17. Using the variance of 17.256 in the Variance Ratio Test gives an F=1.974, which has a p>0.05 (note that we should now use DF=12 for both the numerator and denominator). This indicates that this single datum is having a big effect on the test. We cannot just throw out this data point - that would be inappropriate. However, we can discuss our suspicion as we interpret our conclusion. The conclusion that the variances are different is weak - it may be an artifact of the single datum. There may be nothing of biological importance.

Two-sample *t*-test:

We now wish to answer the question: Is there a difference in the blood glucose level between the two months. We will apply the Two-sample *t*-test. The formula for the *t* statistic given below is not found in your textbook, but Zar provides an equivalent procedure (4th edition: pages 122-125; 5th edition pages 130-134).

Ho: $\mu_1=\mu_2$

$$t=\frac{\overline{X}_1-\overline{X}_2}{\sqrt{\left[\frac{s_1^2(n_1-1)+s_2^2(n_2-1)}{(n_1+n_2-2)}\right]\left[\frac{n_1+n_2}{n_1n_2}\right]}}$$

$$t=\frac{91.1-94.1}{\sqrt{\left[\frac{40.07(14-1)+8.76(13-1)}{(14+13-2)}\right]\left[\frac{14+13}{14\times13}\right]}}=-1.56$$

The DF for this test is DF = $n_1 + n_2 - 2$. For our example, DF = 14 + 13 - 2 = 25. The *t* table in Zar (4th edition: pg. App19; 5th edition: pg.678) indicates that $\alpha(2)$ for $|t|=1.56$ is >0.05 (i.e., p>0.05), and we accept the null hypothesis. There is no significant difference in mean blood glucose levels between January and February.

Remember that the assumptions of the Two-sample *t*-test are RSNDP and equal variances (homoscedasticity). The results of our Variance Ratio Test give us concern about the homoscedasticity assumption. If we apply the "Welch's approximate *t*" (also known as the Behrens-Fisher Test) discussed in Zar, we accept the null hypothesis just as we did above. In fact, the alpha probability of both procedures is virtually identical (i.e. about 0.13). This shows the robustness of the Two-sample *t*-test. The outlier does not have any great effect here.

MANN-WHITNEY U TEST

WHAT IS TESTED
A difference in central tendency between two populations. This does NOT test the means of the samples.

DATA SCALE
Ordinal

NULL HYPOTHESIS
Ho: No difference in central tendency between the two populations. Do NOT use μ or the word "mean" in the null, the means are not being tested here.

One-tailed hypotheses may be tested.

ASSUMPTIONS WITH RESPECT TO DATA
None

DEGREES OF FREEDOM
The sizes of the two samples (n_1 and n_2) are used in looking up the critical values. No degrees of freedom are calculated.

ZAR CHAPTER
8

TYPE
Nonparametric. Parametric analogue is the Two-sample *t*-test.

COMMENTS
This test is about 95.5% as powerful as the Two-sample *t*-test, i.e. it is not able to detect as small a difference between the two samples.

When ranking the data, consider the two samples as one large group for the purpose of ranking. Be careful not to confuse this ranking procedure with the procedure used for the Spearman Rank Correlation.

When doing a two-tailed test, always calculate U and U' and use the larger of the two as your test statistic. Be sure to indicate which one you are using as the test statistic.

As always, be careful of one-tailed tests! When performing a one-tailed test, follow the procedure described in Zar (4^{th} edition: page 149, Table 8.2; 5^{th} edition: page 166, Table 8.2) very carefully. To summarize this procedure:

	Ho: Population 1 ≥ Population 2	Ho: Population 1 ≤ Population 2
Ranking done Low to High	U	U'
Ranking done High to Low	U'	U

Mann-Whitney U Test - Example

The same data used for the example of the Two-sample *t*-test are now analyzed by the Mann-Whitney procedure. Below are the data (blood glucose in mg/dl for January and February) and the ranks. The data have been ranked from low to high, but could have been ranked from high to low in this analysis.

Ho: No difference in central tendency between January and February in blood glucose level.

January	Ranks	February	Ranks
74	1	97	24
97	24	94	15
95	18.5	92	9.5
93	12.5	89	5
85	2.5	90	6
85	2.5	95	18.5
91	7	92	9.5
92	9.5	94	15
95	18.5	94	15
88	4	93	12.5
97	24	99	27
92	9.5	96	21.5
95	18.5	98	26
96	21.5		
n_1=14	R_1=173.5	n_2=13	R_2=204.5

R_i is the sum of the ranks in group i. n_i is the sample size in group i.

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U = 14 \times 13 + \frac{14(14+1)}{2} - 173.5 = 113.5$$

$$U' = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

$$U' = 14 \times 13 + \frac{13(13+1)}{2} - 204.5 = 68.5$$

This is a two-tailed test, therefore the larger of U and U' becomes our test statistic. U = 113.5 is the test statistic.

Examination of Table B.11 in Zar (4^{th} edition: page App96; 5^{th} edition: page 754) indicates that the critical value ($\alpha(2) = .05$) for $n_1 = 13$ and $n_2 = 14$ is 132 (remember that for the purposes of looking in the table, we consider the smaller of the two n values as n_1). Our value of U = 113.5 therefore has an alpha p>0.05. We accept the null hypothesis. We have not found a difference in blood glucose level between the two months.

Shortcut! If you have calculated U, there is a faster way to calculate U’:

$$U' = n_1 n_2 - U$$

In our example, 68.5 = 14 × 13 - 113.5

PAIRED-SAMPLE *t*-TEST

WHAT IS TESTED
If the mean difference between the paired observations of two populations is 0. This is comparable to testing for difference between the means of paired populations, but the null hypothesis must be stated in terms of the mean difference between the paired observations.

DATA SCALE
Ratio-Interval

NULL HYPOTHESIS
Ho: $\mu_d = 0$

One-tailed hypotheses may be tested.

Note that the null must be stated in terms of μ_d.

ASSUMPTIONS WITH RESPECT TO DATA
The differences are a RSNDP of differences. There is NO assumption about the actual data, only about the differences between the paired observations.

DEGREES OF FREEDOM
DF = $n_d - 1$, where n_d is the number of difference values.

ZAR CHAPTER
9

TYPE
Parametric. Nonparametric analogue is the Wilcoxon Paired-Sample Test.

COMMENTS
This is a robust analysis.

Two samples are paired if each observation in one sample is biologically associated with a particular observation in the other sample. The ultimate decision on whether samples are paired depends on the biology of the situation.

Note that this test is really the One-sample *t*-test. It tests the mean of the sample of differences against a constant of 0.

One could test to see if the samples are correlated (using statistical procedures covered near the end of the quarter), and if so, this would tend to indicate that they may be paired. Correlation is a topic that is covered in the second half of the course.

Paired-sample *t*-test - Example

In order to investigate circadian patterns in blood sugar, blood glucose levels are measured in one person at 8:00 am and 8:00 pm on 27 days in January and February. We wish to test for a difference in the levels between the morning (AM) and evening (PM). Data are in mg/dl.

Ho: $\mu_d = 0$

Date	AM	PM	Differences
Jan 14	74	101	-27
Jan 16	97	107	-10
Jan 17	95	97	-2
Jan 18	93	92	1
Jan 19	85	96	-11
Jan 21	85	119	-34
Jan 23	91	109	-18
Jan 24	92	109	-17
Jan 25	95	100	-5
Jan 26	88	152	-64
Jan 27	97	122	-25
Jan 28	92	132	-40
Jan 30	95	111	-16
Jan 31	96	87	9
Feb 1	97	93	4
Feb 2	94	125	-31
Feb 3	92	93	-1
Feb 4	89	115	-26
Feb 6	90	100	-10
Feb 7	95	110	-15
Feb 8	92	93	-1
Feb 10	94	101	-7
Feb 11	94	108	-14
Feb 13	93	91	2
Feb 14	99	110	-11
Feb 15	96	128	-32
Feb 16	98	96	2
Average			-14.78
Standard Deviation			16.277
Standard Error			3.1326

Notice that all of the above statistics are calculated on the Differences column. The analysis is done on the differences between the paired data points. In the formulae below, the subscript "d" refers to "differences".

$$t = \frac{\overline{X}_d}{s_{\overline{X}_d}} = \frac{-14.78}{3.1326} = -4.718 \qquad DF = n_d - 1 = 27 - 1 = 26$$

Again, notice that the sample size (n_d) is the number of differences (27), not the total number of data points (54). Examination of Zar table B.3 (4th edition: page App19; 5th edition: page 678) indicates that the two-tailed alpha probability of -4.718 with 26 degrees of freedom is less than 0.001. Therefore, $p<0.05$ and we reject the null hypothesis. There is a significant difference between the morning and evening in blood glucose level.

WILCOXON PAIRED-SAMPLE TEST

WHAT IS TESTED
A difference in central tendency between paired populations.

DATA SCALE
Ordinal

NULL HYPOTHESIS
Ho: No difference between the paired populations. Do NOT use μ or the word "mean" in the null, no means are being tested.

One-tailed hypotheses may be tested.

ASSUMPTIONS WITH RESPECT TO DATA
The population being sampled is assumed to be symmetrical about the median. This is NOT an assumption of normality. Although the normal distribution is symmetrical about the median, many other distributions also have this property. Know this assumption for closed book questions, but don't worry about whether the data meet this assumption when you are considering using this procedure on a problem.

DEGREES OF FREEDOM
The sample size n_d, where n_d is the number of differences, is used to look up critical values. No degrees of freedom quantities are calculated.

ZAR CHAPTER
9

TYPE
Nonparametric. Parametric analogue is the Paired-sample *t*-test.

COMMENTS
This test is not as powerful as its parametric analogue.

You must rank the absolute value of the differences from low to high. Differences of zero (0) are excluded, i.e. they are not ranked and the sample size (n) is reduced accordingly.

Be careful reading the table (Table B.12, 4th edition: pages App101 - App102, 5th edition: pages 758 - 759). This is the only inferential statistic in this course where the alpha probability decreases as values of the statistic get smaller. In all the other tables, the alpha probability decreases as the values of the statistic get larger.

Don't confuse the Wilcoxon T with Student's *t*. They are very different statistics with very different sampling distributions. When writing the letter, you must use the proper case!

When doing a two-tailed test, calculate both T_+ and T_- and use the smaller of the two as your test statistic. Be sure to indicate which one you are using.

When doing a one-tailed test, follow the procedure found in Zar (4th edition: page 166; 5th edition: page 186) exactly. These can be very tricky. In summary:

Assuming that the differences are calculated as Sample 1 - Sample 2:

For Ho: Population 1 ≤ Population 2; use T_- as the calculated T value.
For Ho: Population 1 ≥ Population 2; use T_+ as the calculated T value.

Wilcoxon Paired-sample Test - Example

We will analyze the blood glucose data from the Paired-sample *t*-test example by the Wilcoxon procedure. Note that we are answering the same biological question, but using a different statistical procedure.

Ho: No difference between the AM and PM in blood glucose levels.

Date	AM	PM	Differences	Rank of |Differences|	Positive Ranks	Negative Ranks
Jan 14	74	101	-27	22		22
Jan 16	97	107	-10	11.5		11.5
Jan 17	95	97	-2	5		5
Jan 18	93	92	1	2	2	
Jan 19	85	96	-11	13.5		13.5
Jan 21	85	119	-34	25		25
Jan 23	91	109	-18	19		19
Jan 24	92	109	-17	18		18
Jan 25	95	100	-5	8		8
Jan 26	88	152	-64	27		27
Jan 27	97	122	-25	20		20
Jan 28	92	132	-40	26		26
Jan 30	95	111	-16	17		17
Jan 31	96	87	9	10	10	
Feb 1	97	93	4	7	7	
Feb 2	94	125	-31	23		23
Feb 3	92	93	-1	2		2
Feb 4	89	115	-26	21		21
Feb 6	90	100	-10	11.5		11.5
Feb 7	95	110	-15	16		16
Feb 8	92	93	-1	2		2
Feb 10	94	101	-7	9		9
Feb 11	94	108	-14	15		15
Feb 13	93	91	2	5	5	
Feb 14	99	110	-11	13.5		13.5
Feb 15	96	128	-32	24		24
Feb 16	98	96	2	5	5	
					$T_+ = 29$	$T_- = 349$

T_+ (sum all of the positive ranks) = 2 + 10 + 7 + 5 + 5 = 29
T_- (sum the absolute values of all of the negative ranks) = 349

This is a two-tailed test, therefore we use the smaller of the two values. Our test statistic is T = 29.

The critical value from Zar Table B.12 (4^{th} edition: pages App101 - App102, 5^{th} edition: pages 758 - 759) with $n_d = 27$ is 107 (note that the sample size n_d is the number of differences, i.e. 27, not the total number of data points (54)). Our calculated value has an alpha probability much less than .05, therefore we reject the null hypothesis. There is a difference between the AM and PM blood glucose levels.

Note that in performing this test, you MUST rank low to high, i.e. the smallest difference gets rank 1, etc. Be very careful in performing a one-tailed Wilcoxon paired-sample test. Follow the procedure found in Zar (4^{th} edition: page 166; 5^{th} edition: page 186) very carefully.

BARTLETT'S TEST

WHAT IS TESTED
Equality of 3 or more population variances (homoscedasticity).

DATA SCALE
Ratio-Interval

NULL HYPOTHESIS
Ho: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 ..... = \sigma_k^2$

ASSUMPTIONS WITH RESPECT TO DATA
All samples are assumed to be RSNDP

DEGREES OF FREEDOM
(you don't need to know this; it is given in Zar, page 202)

ZAR CHAPTER
10

TYPE
Parametric. There is no nonparametric analogue.

COMMENTS
Not a robust analysis. Severely and adversely affected by nonnormal samples.

You do not need to know how to do this analysis, only know the information presented on this sheet.

Homoscedasticity is one of the assumptions of analysis of variance. Bartlett's Test therefore allows a test of that assumption.

As a general rule, Bartlett's Test should only be used with 3 or more groups. If there are only 2 groups, use the Variance Ratio Test. As discussed in Zar (4^{th} edition: page 204; 5^{th} edition: page 221), the actual situation is somewhat more complicated, but we will follow this general rule for the class.

Levene's Test is an analysis which tests the same null hypothesis. Levene's test uses the absolute value (or the square) of the deviation between each datum and its group median. These deviations are then subjected to the one-way ANOVA procedure (see following pages).

ONE-FACTOR (ONE-WAY) ANALYSIS OF VARIANCE (ANOVA)

WHAT IS TESTED
Equality of 2 or more population means

DATA SCALE
Ratio-Interval

NULL HYPOTHESIS
Ho: $\mu_1 = \mu_2 = ... = \mu_k$

ASSUMPTIONS WITH RESPECT TO DATA
All samples are RSNDP. All samples come from populations with equal variances (homoscedasticity).

DEGREES OF FREEDOM
Numerator degrees of freedom is k-1. Denominator degrees of freedom is N-k, where k is the number of groups and N is the total number of data points in all groups combined.

ZAR CHAPTER
10

TYPE
Parametric. Nonparametric analogue is the Kruskal-Wallis Test.

COMMENTS
Always look up the alpha probability of any F value calculated in any ANOVA technique as a one-tailed value.

Know what the quantities TOTAL SS, GROUPS SS, and ERROR SS (and their respective MS quantities) are measuring and how they are calculated. On exams you will be given "puzzles" to solve, and the solutions require you to know these quantities.

Carefully review your lecture notes on the subject of partitioning total variance into between- and within-group components. An understanding of partitioning is critical for understanding ANOVA and all of its related analyses.

Terminology:
Response Variable = **Dependent Variable** = the variable measured by each data point. In the example on the next page, the response variable is infant birth weight. The units of the data will indicate the response variable.
Factor = Grouping Variable = a categorical variable used to place each datum into a particular group. The factor is the independent variable. In the example on the next page, the factor is smoking status of the mother.
Level = category of the factor. In the example on the next page, the factor has three levels, i.e. nonsmoking, 1 pack/day, and 1+ pack/day.
Cell = within the level. The data are in cells. There are three cells in the example on the next page. In the One-factor ANOVA, the cells and the levels are the same thing. This will not be true in Two-factor ANOVA.

The null hypothesis Ho: $\mu_1 = \mu_2$ can be tested either by ANOVA or the Two-sample *t*-test. They are entirely homologous tests in this situation; in fact, the F calculated in ANOVA on a set of data will be the square of a *t* value calculated on the same data. The alpha probabilities of the F and *t* will be exactly the same. One-tailed null hypotheses (Ho: $\mu_1 \leq \mu_2$ or Ho: $\mu_1 \geq \mu_2$) may only be tested with the *t*-test; i.e. you may not use the ANOVA procedure with these hypotheses.

ONE-FACTOR (ONE-WAY) ANOVA - Example

Birth weights (grams) of infants grouped by smoking status of mother. We want to know if smoking status affects birth weight. Birth weight is an indicator of the general health of a newborn infant.

Nonsmoking	1 Pack/Day	1+ Pack/Day
3515	3444	2608
3420	3827	2509
3175	3884	3600
3586	3515	1730
3232	3416	3175
3884	3742	3459
3856	3062	3288
3941	3076	2920
3232	2835	3020
4054	2750	2778
3459	3460	2466
3998	3340	3260
Mean 3613	3363	2901
Standard Deviation 321	369	521

Grand Mean = 3292 (mean of 36 data points)

Ho: $\mu_1 = \mu_2 = \mu_3$

ANOVA Table

Source	SS	DF	MS
Total	8747373	35	249924.9
Groups	3127499	2	1563749.5
Error	5619874	33	170299.2

$$F = \frac{Groups\ MS}{Error\ MS} = \frac{1563749.5}{170299.2} = 9.18$$

$DF_{numerator} = 2$, $DF_{denominator} = 33$

See Zar, 4th edition: page App22; 5th edition page 681) $p<0.05$ Reject Ho:

Note: Groups are homoscedastic as indicated by Levene's Test.
Each group is normally distributed as indicated by the Shapiro-Wilks Test.
Therefore, these data meet both assumptions of ANOVA.

Note: The above ANOVA table was generated by a computer. In the following pages, the calculations for values in the table are shown. However, the actual numbers obtained differ from the above table. This is due to rounding error.

Sources of Variation in a One-factor ANOVA

The key to understanding ANOVA (and to answering correctly the ANOVA problems on the exams) is to understand the sources of variation. Let's explore this topic.

First, review some terminology from basic statistics:

Variability = **Variation** the property of being different. The opposite of a constant. The baby weights exhibit variability (there is variation in birth weight; not all babies weigh the same at birth). These terms do not refer to unique statistical quantities, but just to the general property of being different.

Sum of Squares (SS) the sum of squared deviations. This is an estimate of variability. Sums of squares are always affected by the sample sizes.

Variance a sum of squares divided by the appropriate degree of freedom. The degree of freedom reflects the sample size, so variances are not affected by sample sizes.

Standard Deviation is the positive square root of a variance. The variance and standard deviation are estimates (statistics) of variability. The standard deviation can be more easily interpreted, because it is in the same units as the original data.

Mean Square (MS) a variance used in an ANOVA technique. Notice that since an MS is a variance, that an MS must equal a SS/DF. The V in ANOVA refers to the MS quantities used in the analysis.

Second, let's define some symbols we'll be using:

X_{ij} datum in row i, column j. Since our columns are our groups (i.e. our levels), the value for j indicates what group (level) the datum belongs to. In our example, the value for j indicates which smoking group the baby is in. The i value has no special meaning, it just allows us to refer to individual babies within a smoking group.

$\overline{X}$ the grand mean, i.e. the mean of all the data points. The grand mean for all 36 babies is 3292 g.

$\overline{X}_j$ the mean for group (level) j. These are the means of the smoking groups, i.e. the means we are testing. The mean for the nonsmoking babies (j=1) is 3613; the mean for the 1 Pack/Day group (j=2) is 3363; and the mean for the 1+ Pack/Day group (j=3) is 2901. These means are the estimates of the population means we see in the null hypothesis Ho: $\mu_1 = \mu_2 = \mu_3$.

n_j the number of data points in group j. Each of our groups has an n_j value of 12. This is a *balanced* design. If the sample sizes were different in the groups, we would have an *unbalanced* design. While it is usually desirable to have a balanced design, an unbalanced design would not be analyzed any differently (note: this is not the case when we get to Two-factor ANOVA later in the course).

N the total number of data points, i.e. in all groups combined. In our baby weight example, there are 36 babies, therefore N=36.

k the number of groups or levels. In the baby example, k=3.

Now, finally, the sources of variation in our One-factor ANOVA:

TOTAL is the variability of *all* the data points. It is the variability of all the data points from the grand mean.

GROUPS is the variability of the group (level) means. It is a measure of how far away from one another the group means are. Notice that if the group means are far away from one another, we are more likely to reject the null hypothesis that the means are equal.

Groups variability is also thought of as the effect of the treatment. For example, if smoking affects baby birth weight, then the means of the smoking groups should be far apart, i.e. groups variability will be large.

Measuring groups variability is accomplished by using the grand mean as a central reference point, i.e. we calculate the distance between each group mean and the grand mean. The rationale here is that if the group means are close together, they will all be close to the grand mean. However, if the group means are far apart, then at least some of the group means have to be pretty far from the grand mean.

ERROR is the variability *within* the smoking groups. It does not measure the effect of smoking, because within a smoking group, each baby has received the same smoking "treatment".

Error measures variability caused by everything else other than smoking. We call this "common" or "unexplained" variability, i.e. it is not explained by our data set. We cannot explain why there is variability within a smoking group - it is due to factors we haven't considered. Error variability gives us a measure of how accurate our group mean estimates are. In other words, how much *error* is involved in estimating the means of the smoking groups.

In the homoscedasticity assumption, we are assuming that the within group variability is the same for all groups. This allows us to calculate a single, pooled estimate of within group variability. This single, pooled estimate is Error variability.

Now, we'll look at partitioning of variability.

In partitioning, we take TOTAL variability and partition it into GROUPS and ERROR. In other words, TOTAL = GROUPS + ERROR.

Why do we partition? So that we can test the null hypothesis! Remember, our null is that the population means are equal. To test this null, we need to know (1) how far apart the group means are from one another; and (2) how much error is involved in estimating the means of the smoking groups. (1) is GROUPS variability, and (2) is ERROR. Partitioning is how we go about getting these two critical quantities.

The actual, mathematical partitioning is done with the Sums of Square (SS) quantities. In other words, TOTAL SS = GROUPS SS + ERROR SS. Let's see how we calculate these SS quantities. Don't forget the definitions of these sources of variation as you look at the details of doing the calculations. The most important thing for you to learn is the concept of the sources of variation. The calculations are presented here just to help you understand the quantities. In "real life", the calculations are done by a computer. But the computer doesn't understand what it has calculated; that's your job!

$$Total\ SS = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ij} - \overline{X})^2$$

Look at the equation, and remember what TOTAL variability is - it is the variability of all the data points. What this formula says to do is to take each datum, subtract the grand mean, square the difference, and finally, sum all of the squared differences. Remember that summing is done last!

First, calculate the squared deviation of each datum from the grand mean of 3292:

$(3515-3292)^2 = 49729$	$(3444-3292)^2 = 23104$	$(2608-3292)^2 = 467856$
$(3420-3292)^2 = 16384$	$(3827-3292)^2 = 286225$	$(2509-3292)^2 = 613089$
$(3175-3292)^2 = 13689$	$(3884-3292)^2 = 350464$	$(3600-3292)^2 = 94864$
$(3586-3292)^2 = 86436$	$(3515-3292)^2 = 49729$	$(1730-3292)^2 = 2439844$
$(3232-3292)^2 = 3600$	$(3416-3292)^2 = 15376$	$(3175-3292)^2 = 13689$
$(3884-3292)^2 = 350464$	$(3742-3292)^2 = 202500$	$(3459-3292)^2 = 27889$
$(3856-3292)^2 = 318096$	$(3062-3292)^2 = 52900$	$(3288-3292)^2 = 16$
$(3941-3292)^2 = 421201$	$(3076-3292)^2 = 46656$	$(2920-3292)^2 = 138384$
$(3232-3292)^2 = 3600$	$(2835-3292)^2 = 208849$	$(3020-3292)^2 = 73984$
$(4054-3292)^2 = 580644$	$(2750-3292)^2 = 293764$	$(2778-3292)^2 = 264196$
$(3459-3292)^2 = 27889$	$(3460-3292)^2 = 28224$	$(2466-3292)^2 = 682276$
$(3998-3292)^2 = 498436$	$(3340-3292)^2 = 2304$	$(3260-3292)^2 = 1024$

Next, sum the squared deviations:

49729 + 23104 + 467856 + 16384 + 286225 + 613089 + 13689 + 350464 + 94864 +
86436 + 49729 + 2439844 + 3600 + 15376 + 13689 + 350464 + 202500 + 27889 +
318096 + 52900 + 16 + 421201 + 46656 + 138384 + 3600 + 208849 + 73984 +
580644 + 293764 + 26416 + 27889 + 28224 + 682276 + 498436 + 2304 + 1024 = 8747374

Therefore, Total SS = 8747374

$$Groups\ SS = \sum_{j=1}^{k} n_j\,(\overline{X}_j - \overline{X})^2$$

Look at the equation and remember what GROUPS variability is - it is a measure of how far apart the group means are from each other. The formula says to calculate the squared distance of each group mean from the grand mean, and then multiply by the sample size in that group. The last step is to sum across all the groups. For our baby data, the calculation is done like this:

Nonsmoking Group: $12(3613-3292)^2 = 12 \times 103041 = 1236492$
1 Pack/Day Group: $12(3363-3292)^2 = 12 \times 5041 = 60492$
1+Pack/Day Group: $12(2901-3292)^2 = 12 \times 152881 = 1834572$

3131556 = Groups SS

Now go to the next page to examine the calculation for Error SS.

$$Error\ SS = \sum_{j=1}^{k}\left[\sum_{i=1}^{n_j}(X_{ij} - \overline{X}_j)^2\right]$$

Look at the equation and remember what ERROR variability is - it is a measure of pooled, *within group* variability. The formula says to calculate a SS for each group (i.e. each datum minus its group mean); and then to sum the SS across all the groups. This is more confusing than difficult. Let's see how it's done for the baby weight data.

First, calculate a SS for each group:

Nonsmoking	1 Pack/Day	1+ Pack/Day
$(3515-3613)^2$= 9604	$(3444-3363)^2$= 6561	$(2608-2901)^2$= 85489
$(3420-3613)^2$= 37249	$(3827-3363)^2$=215296	$(2509-2901)^2$= 153664
$(3175-3613)^2$=191844	$(3884-3363)^2$=271441	$(3600-2901)^2$= 488601
$(3586-3613)^2$= 729	$(3515-3363)^2$= 23104	$(1730-2901)^2$=1371241
$(3232-3613)^2$=145161	$(3416-3363)^2$= 2809	$(3175-2901)^2$= 75076
$(3884-3613)^2$= 73441	$(3742-3363)^2$=143641	$(3459-2901)^2$= 311364
$(3856-3613)^2$= 59049	$(3062-3363)^2$= 90601	$(3288-2901)^2$= 149769
$(3941-3613)^2$=107584	$(3076-3363)^2$= 82369	$(2920-2901)^2$= 361
$(3232-3613)^2$=145161	$(2835-3363)^2$=278784	$(3020-2901)^2$= 14161
$(4054-3613)^2$=194481	$(2750-3363)^2$=375769	$(2778-2901)^2$= 15129
$(3459-3613)^2$= 23716	$(3460-3363)^2$= 9409	$(2466-2901)^2$= 189225
$(3998-3613)^2$=148225	$(3340-3363)^2$= 529	$(3260-2901)^2$= 128881
1136244	1500313	2983321

Now that we have a within group SS for each group, we sum them to calculate Error SS. In statistical terms, we are *pooling* the SS of the groups.

Nonsmoking Group SS	1136244
1 Pack/Day Group SS	1500313
1+ Pack/Day Group SS	2983321
Error SS =	5619878

Now we have our SS quantities (Total SS, Groups SS, and Error SS). **BUT, WAIT!!!** I thought that Total SS = Groups SS + Error SS!! If you look at our quantities, 8747374 ≠ 3131556 + 5619878!! In fact, 3131556 + 5619878 = 8751434!! What's up with this? What's happening? The answer is simple: it's due to **rounding error**.

We used 3292 for the grand mean in our calculations, but that's rounded to the nearest gram. To ten decimal places the grand mean is 3292.1111111111. We also used group means rounded to the nearest gram. To ten decimal places, the group means are: Nonsmoking = 3612.6666666667, 1 Pack/Day = 3362.5833333333, and 1+ Pack/Day = 2901.0833333333.

If you redo all of the above calculations using the means to ten decimal places, you will get values that are much more accurate and much closer to adding up (i.e. having Total SS = Groups SS + Error SS). You'll also go crazy punching all of those numbers on your calculator. This is a good example of why, in "real life", statistical analyses are always done on a computer. A computer would typically do these calculations to at least 16 decimal places.

If you look back several pages at the ANOVA table for the baby weights, you will see that the SS values are different from what we have calculated here. This again is due to rounding error; the ANOVA table was generated by a computer.

A Graphical Approach to Sources of Variation

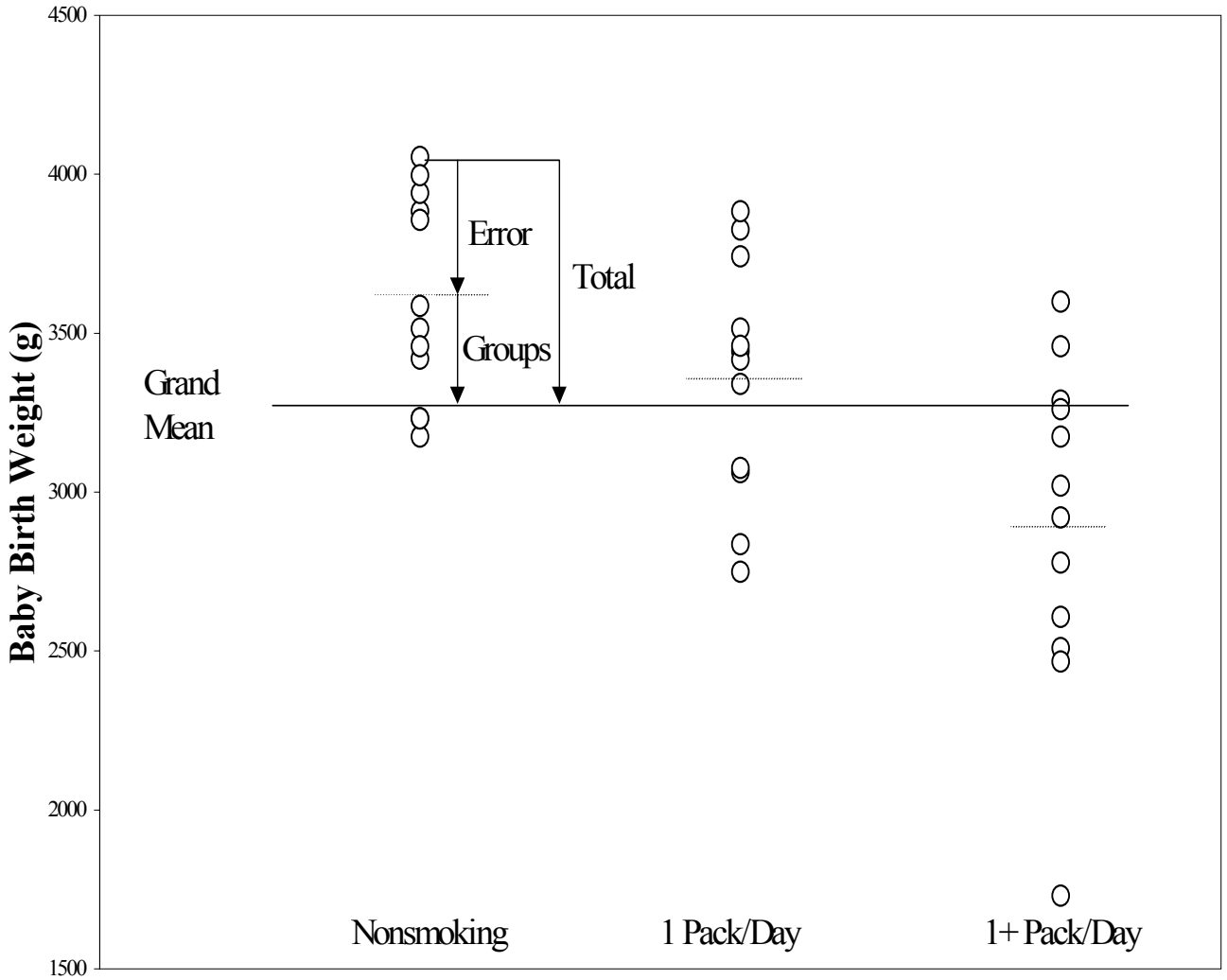
The graph below shows each baby graphed by weight as a function of smoking group. The solid line is the grand mean (3292), while the shorter, dashed lines represent the group means (3613 for nonsmoking, 3363 for 1 pack/day, and 2901 for 1+ pack/day).

Total variation is the squared distance between each baby (open circle) and the grand mean.

Groups variation is the squared distance between the group means and the grand mean.

Error variation is the squared distance between each baby and the mean of its smoking group.

Total, Groups, and Error distances are shown for one baby. This is a "graphical representation" of the partitioning of Total into Groups and Error. Note that this is not a mathematical representation, i.e. it is the sum of the squared distances that are additive (Total SS = Groups SS + Error SS), not the actual distances as the graph might lead you to believe.



A shortcut method of calculating Error SS:

Several pages previously, when we first saw the infant birth weight data, we were given the means and the standard deviations of the smoking groups. Look back to that page (it's the one with the ANOVA table) now.

We can use those standard deviations to calculate Error SS. If we square the standard deviations, we get the variances of the groups (remember your basic statistics!).

Variance of Nonsmoking group = 321^2 = 103041
Variance of 1 Pack/Day group = 369^2 = 136161
Variance of 1+ Pack/Day group = 521^2 = 271441

Since the Variance = SS/DF, it must be true that SS = Variance × DF. Since these are sample variances (see the cover of this Test Pac), the DF value must be n-1, where n=12 for each group.

So to calculate the SS for each group, we multiply the variance × 11:

Nonsmoking SS = Variance × DF = 103041 × 11 = 1133451
1 Pack/Day SS = Variance × DF = 136161 × 11 = 1497771
1+Pack/Day SS = Variance × DF = 271441 × 11 = 2985851

Remember that Error SS is the pooled, within group SS. Since we have the within group SS for each group, all we need to do is pool them, i.e. sum them:

1133451 + 1497771 + 2985851 = 5617073 = Error SS

You should notice that this is not the same value we got before (which was 5619878). Can you guess why? That's right! It's simply due to rounding error!

Degrees of Freedom and Mean Squares

We wish to compare our sources of variation, specifically we need to compare Groups and Error. We can't compare SS quantities because they are affected by sample size. We must convert each SS to a variance by dividing by the appropriate degree of freedom (DF). A variance (SS/DF) used in an ANOVA technique is called a **Mean Square**. Let's look at each source of variation:

Look at our previous calculation for Total SS. We summed 36 squared deviations, so clearly the sample size N affects the magnitude of the number. **The DF for Total is calculated as N-1.**
Total MS = Total SS / Total DF = 8747374 / 35 = 249924.9

When we calculated Groups SS, we squared the difference between each group mean and the grand mean. So Groups SS is affected by k, the number of groups. **Groups DF is calculated as k-1.**
Groups MS = Groups SS / Groups DF = 3131556 / 2 = 1565778

For Error SS, we had to calculate a SS for each group, and then pool (sum across all the groups). Therefore, Error SS is affected by the number of data points (N) as well as the number of groups (k). **Error DF is calculated as N-k.**
Error MS = Error SS / Error DF = 5619878 / 33 = 170299.3

Error DF is also a pooled value. The DF within each group is n_j-1. If you sum the n_j-1 for each group, you get N-k. For our example: n_1-1 + n_2-1 + n_3-1 = n_1+n_2+n_3-1-1-1 = N-k. (12-1)+(12-1)+(12-1) = 36 - 3 = 33.

Error MS can also be calculated as the mean of the variances of the three groups: (103041 + 136161+ 271441)/3 = 170214 (Difference from above due to rounding error) The above calculation works because the sample sizes for all three groups is the same (n_j = 12). If the sample sizes were different, a weighted mean would have to be calculated (weighted by sample size).

Mean Squares and the Test of the Null Hypothesis

Now that we have our Mean Squares, we are ready to test the null hypothesis (for the baby weight data, Ho: $\mu_1 = \mu_2 = \mu_3$). We will reject the null if the means are far apart relative to the amount of error involved in estimating them. In other words, if there is an effect of smoking on birth weight, the means of the smoking groups should be far apart. They should be further apart than would be expected just by random sampling error (i.e. the error involved in estimating the means).

Groups MS is a variance measuring how far apart the means of the smoking groups are from each other, i.e. the effect of smoking.

Error MS is a variance measuring the random sampling error involved in estimating the means.

So, **we will reject the null if Groups MS is significantly greater than Error MS.**

We perform an F test by calculating $F = \frac{Groups\ MS}{Error\ MS} = \frac{1565778}{170299.3} = 9.194$

The DF for the numerator is Groups DF, i.e. k-1 (3-1=2).

The DF for the denominator is Error DF, i.e. N-k (36-3=33).

Because we will reject the null if Groups MS is significantly greater than Error MS, this F statistic is considered to be one-tailed (remember, the F ratio in all ANOVA techniques is always one-tailed; there are no exceptions).

We look at the F table in Zar (Table B.4) for the appropriate degrees of freedom (4^{th} edition: page App22; 5^{th} edition: page 681). You will notice that there is no denominator DF of 33, so we'll look at the row where Denom. DF is 30. We see that the one-tailed alpha probability of an F=9.194 with 2,30 degrees of freedom is $0.001>p>0.0005$. So, with our larger Denom. DF, it is true that $p<0.05$, and we reject Ho:.

Smoking status does appear to affect infant birth weight. The magnitude of this effect can be quantified by calculating the explained variation (Groups SS) as a proportion of the total variation (Total SS).

$$\frac{\text{Groups SS}}{\text{Total SS}} = \frac{3127499}{8747373} = 35.7\%$$

Smoking status explains about 36% of the variability in birth weight, therefore, about 64% is unexplained.

You should work very hard to gain an understanding of the sources of variation in ANOVA. These concepts are critically important, both for the exams and for your career as a biologist. The details of doing all of the calculations "from scratch" are provided to help you understand the concepts. You will not have to do all of these calculations on the exams, you *will* have to understand the concepts. See the practice exams for typical questions.

A note about the formulas in Zar.....
If you look at Zar (4^{th} edition: Example 10.1, pages 180-181; 5^{th} edition: Example 10.1b, page 196) you will see different formulas to calculate the SS quantities in the ANOVA. These formulas are equivalent to the formulas used here in the Test Pac. The formulas in Zar are called "machine formulas". They produce less rounding error (and you've seen how important that is!), and are somewhat faster if you're doing an ANOVA from scratch using a calculator - something you should never do! In "real life", ANOVA should only be done on a computer. The probability of making a mistake on a calculator is too high.

MULTIPLE COMPARISONS

WHAT IS TESTED
After the null hypothesis of equality of means is rejected by Analysis of Variance, a multiple comparison test attempts to find which means are different.

DATA SCALE
Ratio-Interval

NULL HYPOTHESIS
(You do not need to know this. It can vary depending on which analysis is used).

ASSUMPTIONS WITH RESPECT TO DATA
All samples RSNDP. All samples from populations with equal variances (homoscedasticity).

DEGREES OF FREEDOM
(You do not need to know this.)

ZAR CHAPTER
11

TYPE
Parametric. Nonparametric methods are available.

COMMENTS
There are many multiple comparison procedures available, and selecting the best one in each situation is not a simple matter. The statistical community has not yet settled on a "best" procedure.

You should know the following names: 1) **Tukey Test**; and 2) **Newman-Keuls** Test. These are commonly used procedures. The **Tukey-Kramer Test** may be used when the sample sizes of the groups are not the same. **Dunnett's Test**, is used in the special situation where you wish to test the mean of a control group against experimental groups, but not test the experimental groups against one another. **Scheffé's Test** can be used for multiple comparisons (although it performs poorly due to lack of power), but also allows you to test combinations of groups against one another - a procedure called *multiple contrasts*. An example of multiple contrasts would be to test mean birth weight for the 1 pack/day and 1+ pack/day groups combined against the mean of the nonsmoking babies.

A multiple comparison procedure should only be used after the null hypothesis in ANOVA has been rejected. This is because the ANOVA is a more powerful test than the multiple comparison procedure. It is possible to reject the null in ANOVA, but have a multiple comparison procedure fail to find any differences among any of the means.

The Tukey Test and the Newman-Keuls Test were performed on the baby birth weight data from the One-factor ANOVA example, and both resulted in the same conclusion: i.e. the nonsmoking babies were not different from the 1 pack/day babies, but both of these groups were different from the 1+ pack/day group. In other words:

Nonsmoking mean (3613) = 1 pack/day mean (3363) > 1+ pack/day mean (2901)

When both tests give you the same result (as above), you can be fairly confident that the result is appropriate. When different tests give you different conclusions, then your situation is complicated. This may be a time to consult a statistician familiar with ANOVA and multiple comparison problems.

KRUSKAL - WALLIS TEST

WHAT IS TESTED
A difference in central tendency among 3 or more populations.

DATA SCALE
Ordinal

NULL HYPOTHESIS
Ho: No difference in central tendency among the populations. Do NOT use μ or the word "mean" in the null; the means are not being tested.

ASSUMPTIONS WITH RESPECT TO DATA
None

DEGREES OF FREEDOM
(You do not need to know this)

ZAR CHAPTER
10

TYPE
Nonparametric. Parametric analogue is the Analysis of Variance.

COMMENTS
You will not be required to perform this test. You are required to know the information presented on this sheet.

The Kruskal-Wallis is about 95.5% as powerful as the ANOVA.

The Kruskal-Wallis should be used when there are three or more samples. If there are only two samples, the Mann-Whitney U is the correct choice. With two samples, the Kruskal-Wallis is actually identical to the Mann-Whitney U, but the Kruskal-Wallis is generally recognized as a multiple sample (i.e. >2) test.

Multiple comparison procedures may be employed after rejection of the null hypothesis. They are discussed in Zar (4^{th} edition: pages 223-226; 5^{th} edition: pages 239-243).

TWO-FACTOR (TWO-WAY) ANALYSIS OF VARIANCE

WHAT IS TESTED
Equality of means among levels of factors, when the data can be categorized by two factors. In one model, the levels of one of the factors are considered blocks, i.e. particular observations in different groups are biologically related to one another (an extension of the paired concept). Interaction between factors may be tested in certain models (data structures).

DATA SCALE
Ratio-Interval

NULL HYPOTHESIS
Several hypotheses may be tested:

1. No difference among the means of the levels of one factor.
2. No difference among the means of the levels of the second factor.
3. No interaction among the levels of the two factors.

The exact null hypotheses which may be tested in any particular problem is affected by the data, the model, and the biological question(s) of interest.

ASSUMPTIONS WITH RESPECT TO DATA
Data within cells are assumed to be RSNDP and homoscedastic. Since sample sizes within cells are usually low, the assumption is sometimes applied to levels of the factors as an approximation.

DEGREES OF FREEDOM
Varies depending on the model and tests used. See Zar.

ZAR CHAPTER
12

TYPE
Parametric. Nonparametric procedures are available. The procedures differ based on the model and hypothesis test being made. See Zar.

COMMENTS
This is a rather complex analysis, but it is very important. Much biological data, especially experimental data (e.g. microbiology or physiology lab data), are analyzed with these techniques.

An important feature of this analysis, for our purposes, is that it allows us to extend the concept of the "paired analysis" to situations where we have more than two samples. Since "pair" refers to two, we use the term "block" when there are three or more samples. This analysis allows us to test the means of the samples by first removing any variability due to differences among blocks. This is the multi-sample extension of the Paired-sample *t*-test. See Zar (4^{th} edition: pages 250-254; 5^{th} edition pages 270-272) for relevant information.

Study and be very familiar with the manner in which total variability is partitioned in various Two-way ANOVAs. Know what among levels, interaction, and error components mean. This analysis is computationally intense, and you won't be asked to do one from scratch. You will be expected to interpret the results of such an analysis, and you will be tested for an understanding of the components of the total variability partition.

In a Two-factor ANOVA, a **cell** is the intersection of the levels of the factors. The data are found in the cells.

In actual practice, the computations for this analysis should be left to the computer.

Two-factor (Two-way) ANOVA - Example: randomized block

Data are weight gains (grams) of guinea pigs on 4 different diets. Each block of four animals was housed in a different room. The light/dark cycle, temperature, and noise level was different in each room. (Note: this is example 12.4 in the 4th edition of Zar; it is not in the 5th edition.)

Ho: Mean weight gain on the 4 diets is the same ($\mu_1 = \mu_2 = \mu_3 = \mu_4$)

	Diet 1	Diet 2	Diet 3	Diet 4	Block Means
Block 1	7.0	5.3	4.9	8.8	6.5
Block 2	9.9	5.7	7.6	8.9	8.025
Block 3	8.5	4.7	5.5	8.1	6.7
Block 4	5.1	3.5	2.8	3.3	3.675
Block 5	10.3	7.7	8.4	9.1	8.875
Diet Means	8.16	5.38	5.84	7.64	
Grand Mean	6.755				

Sources of Variation

Total SS is defined exactly the same as in the One-factor ANOVA, i.e. the sum of the squared deviation of each data point from the grand mean. The calculation would look like this:

$(7.0-6.755)^2=0.06$ $(5.3-6.755)^2=2.12$ $(4.9-6.755)^2=3.44$ $(8.8-6.755)^2=4.18$
$(9.9-6.755)^2=9.89$ $(5.7-6.755)^2=1.11$ $(7.6-6.755)^2=0.71$ $(8.9-6.755)^2=4.60$
$(8.5-6.755)^2=3.05$ $(4.7-6.755)^2=4.22$ $(5.5-6.755)^2=1.58$ $(8.1-6.755)^2=1.81$
$(5.1-6.755)^2=2.74$ $(3.5-6.755)^2=10.60$ $(2.8-6.755)^2=15.64$ $(3.3-6.755)^2=11.94$
$(10.3-6.755)^2=12.57$ $(7.7-6.755)^2=0.89$ $(8.4-6.755)^2=2.71$ $(9.1-6.755)^2=5.50$

$$0.06 + 2.12 + 3.44 + 4.18 + 9.89 + 1.11 + 0.71 + 4.60 + 3.05 + 4.22 + 1.58 + 1.81 + + 2.74 + 10.60 + 15.64 + 11.94 + 12.57 + 0.89 + 2.71 + 5.50 = 99.3495 = \text{Total SS}$$

In a two-factor ANOVA, we have two grouping variables (i.e. two factors). Our two factors are Diets and Blocks. We calculate a "Groups SS" for each factor - we'll call them Diets SS and Blocks SS. Each of the SS is calculated as if that group were the only group in a one-factor ANOVA. In other words, for Diets SS, we take the squared deviation of each Diet mean from the grand mean times the sample size in the Diet. For Blocks SS, we take the squared deviation of each Block mean from the grand mean times the sample size in the Block. It's almost easier to do the calculations than to describe them! Here's how they're done:

Diets SS

$5\times(8.16-6.755)^2=9.870$
$5\times(5.38-6.755)^2=9.453$
$5\times(5.84-6.755)^2=4.186$
$5\times(7.64-6.755)^2=3.916$

Diets SS = 27.4255

Blocks SS

$4\times(6.500-6.755)^2= 0.260$
$4\times(8.025-6.755)^2= 6.452$
$4\times(6.700-6.755)^2= 0.012$
$4\times(3.675-6.755)^2=37.946$
$4\times(8.875-6.755)^2=17.977$

Blocks SS = 62.647

The only way to calculate Error SS in this ANOVA model is by subtraction, i.e.
Error SS (Remainder SS) = Total SS - Diets SS - Blocks SS
Error SS (Remainder SS) = 99.3495 - 27.4255 - 62.647 = 9.277

Zar uses the term Remainder SS rather than Error SS. Remainder SS is the better term because it is calculated by subtraction, and also for statistical reasons that we aren't in a position to deal with at this time. For this class, either Remainder SS or Error SS will be acceptable.

Degrees of Freedom

The Total DF is N-1 (where N is the total number of data points). This is exactly the same as in One-factor ANOVA. In fact, Total DF is always N-1 in all ANOVAs.

Diets and Blocks are our factors. The DF associated with a factor in an ANOVA is always the number of groups (levels) minus 1. The Diets factor has four groups (levels), so the Diets DF = 4 - 1 = 3. The Blocks factor has five groups (levels), so the Blocks DF = 5 - 1 = 4.

Calculate Error DF by subtraction:
Total DF - Diets DF - Blocks DF = 19 - 3 - 4 = 12.

Mean Squares (MS)

As always, Mean Squares are calculated as SS/DF. Mean Squares are variances.

Now we are ready for our ANOVA table:

ANOVA Table

Source	SS	DF	MS
Total	99.3495	19	5.2289
Diets	27.4255	3	9.1418
Blocks	62.647	4	15.6618
Error	9.277	12	0.7731

Test equality of diet means, Ho: $\mu_1 = \mu_2 = \mu_3 = \mu_4$

$$F = \frac{Diets\ MS}{Error\ MS} = \frac{9.1418}{0.7731} = 11.825$$

DF = 3, 12 $p<0.05$ Reject Ho:

Test equality of block means, Ho: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

$$F = \frac{Blocks\ MS}{Error\ MS} = \frac{15.6618}{0.7731} = 20.259$$

DF = 4, 12 $p<0.05$ Reject Ho:

The test of equality of block means is inappropriate if there is interaction between diets and blocks. Because there are no replications, interaction cannot be tested.

See the next page for more discussion of interaction.

Interaction

Interaction is defined as the pattern of change of the response variable across the levels of one factor as a function of the levels of the second factor.

The easiest way to understand interaction is to examine the concept graphically.

The figure below is a graph of the guinea pig weight gain data from the previous page. Note that each line represents one diet (i.e. one level of the Diet factor).

We are interested in the pattern of change in weight gain in the diets as a function of the blocks (rooms). As the graph shows, the diets are variable.

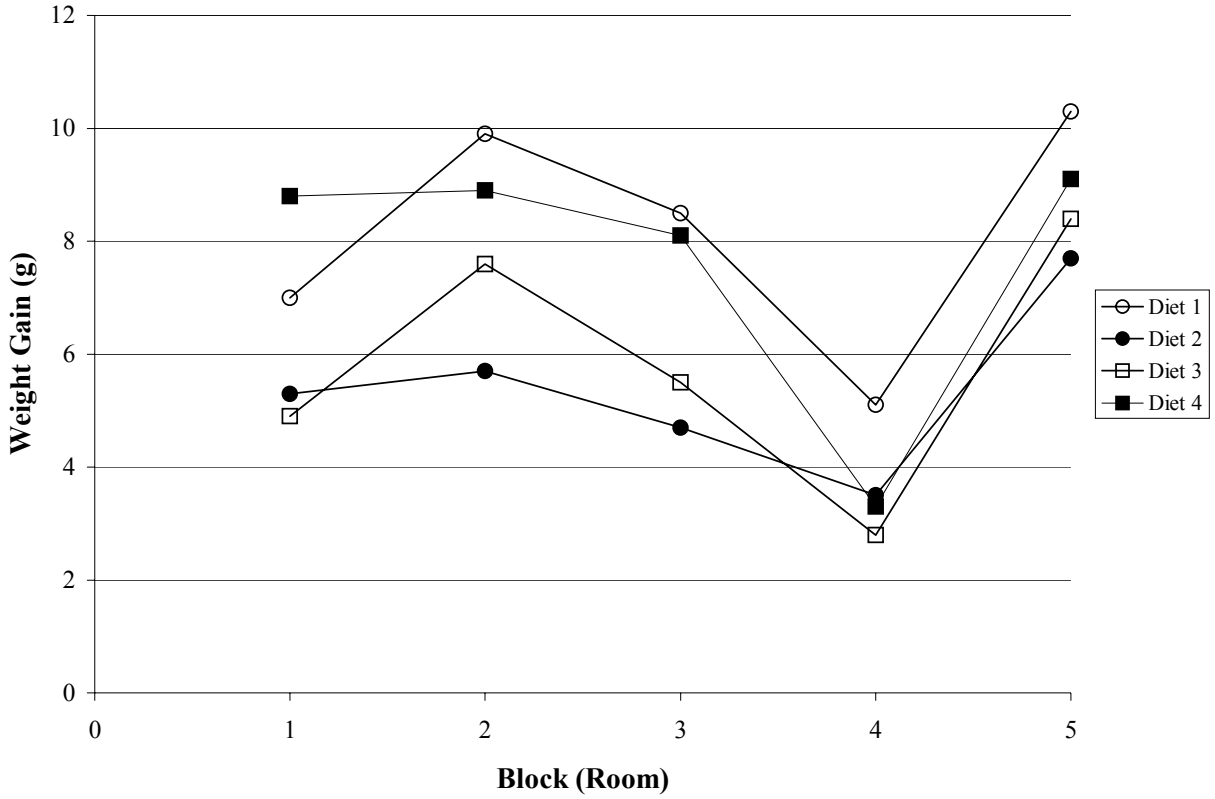
Note that diets 1 and 3 seem to increase from block 1 to block 2, while diets 2 and 4 seem about the same in blocks 1 and 2.

If the diets all respond in the same way to the different blocks (i.e. the four lines are parallel), we say there is no interaction. If the diets respond differently, we say there is interaction.

If diets 1 and 4 really respond differently, there is interaction. On the other hand, if this is just random sampling error, then there is no interaction.

How can we determine whether or not there is interaction? In this data set, we cannot test for interaction because there are no replications in the cells. Because there is only one datum per cell, it is not possible to determine any error (i.e. within cell error) associated with the datum.

In our next example, we will see a two-factor ANOVA with replications in the cells. This replication will allow a test of the null hypothesis Ho: No Interaction.



Two-factor (Two-way) ANOVA - Example: with replications

A clinic that does health evaluations is studying the effect of smoking. The clinic evaluates people using one of two devices: a stationary bicycle and a treadmill. While the subject is on the bike or treadmill, their oxygen consumption is measured, and the time (in minutes) required for the subject to reach their maximum oxygen consumption is noted. A recent experiment involved 18 individuals: 6 nonsmokers, 6 moderate smokers, and 6 heavy smokers. In each smoking group, 3 individuals were randomly assigned to the bike and the other 3 to the treadmill. The data from the experiment are below.

Is there an effect of smoking on time to maximum O_2 consumption?
Do the bicycle and treadmill produce different times (i.e. is there a device effect)?
Is the change in time for the smoking groups independent of what device is used?

	Nonsmokers			Moderate Smokers			Heavy Smokers		
Bicycle	12.8	13.5	11.2	10.9	11.1	9.8	8.7	9.2	9.5
Treadmill	17.8	18.1	16.2	15.5	13.8	16.2	14.7	13.2	10.1

Following are means ± standard deviations with sample sizes in parentheses:

Grand Mean: 12.906 ± 3.0153 (n=18)

Means for the levels of the smoking factor:
Nonsmokers: 14.933 ± 2.8423 (n=6)
Moderate Smokers: 12.883 ± 2.6574 (n=6)
Heavy Smokers: 10.900 ± 2.4519 (n=6)

Means for the levels of the device factor:
Bicycle: 10.744 ± 1.6272 (n=9)
Treadmill: 15.067 ± 2.4829 (n=9)

Cell means:
Bicycle, Nonsmoking: 12.500 ± 1.179 (n=3)
Bicycle, Moderate Smoking: 10.600 ± 0.700 (n=3)
Bicycle, Heavy Smoking: 9.133 ± 0.404 (n=3)
Treadmill, Nonsmoking: 17.367 ± 1.021 (n=3)
Treadmill, Moderate Smoking: 15.167 ± 1.234 (n=3)
Treadmill, Heavy Smoking: 12.667 ± 2.346 (n=3)

ANOVA Table

Source	SS	DF	MS	Ho:	F	p	Conclusion
Total	154.5695	17	9.0923				
Smoking	48.8078	2	24.4039	$\mu_1 = \mu_2 = \mu_3$	14.48	<0.05	Reject Ho:
Device	84.0672	1	84.0672	$\mu_1 = \mu_2$	49.88	<0.05	Reject Ho:
Interaction	1.4678	2	0.7339	No Interaction	0.44	>0.05	Accept Ho:
Error	20.2267	12	1.6856				

There is a smoking effect. We rejected the null that the smoking means were equal.

There is a device effect. Times on the treadmill were significantly longer.

There is no interaction, i.e. the change in the time across the smoking groups is independent of the device used.

Sources of Variation

TOTAL
As always, Total SS is the sum of the squared deviation of each datum from the grand mean:

$(12.8-12.906)^2 = 0.011$ $(10.9-12.906)^2 = 4.024$ $(8.7-12.906)^2 = 17.690$
$(13.5-12.906)^2 = 0.353$ $(11.1-12.906)^2 = 3.262$ $(9.2-12.906)^2 = 13.734$
$(11.2-12.906)^2 = 2.910$ $(9.8-12.906)^2 = 9.647$ $(9.5-12.906)^2 = 11.601$

$(17.8-12.906)^2 = 23.951$ $(15.5-12.906)^2 = 6.729$ $(14.7-12.906)^2 = 3.218$
$(18.1-12.906)^2 = 26.978$ $(13.8-12.906)^2 = 0.799$ $(13.2-12.906)^2 = 0.086$
$(16.2-12.906)^2 = 10.850$ $(16.2-12.906)^2 = 10.850$ $(10.1-12.906)^2 = 7.874$

0.011+0.353+2.910+4.024+3.262+9.647+17.690+13.734+11.601+23.951+26.978+10.850+
6.729+0.799+10.850+3.218+0.086+7.874 = 154.567 = Total SS

Is there a faster way to get this? Look back at the previous page, and notice that we were told that the grand mean was 12.906, and that the standard deviation of all the data points was 3.0153. If we square the standard deviation we get the variance of all the data points, which is Total MS! So, $3.0153^2 = 9.092$ = Total MS. Since Total DF = 17, then Total SS = 9.092 × 17 = 154.56 = Total SS. The key here is to know that if we know the standard deviation of all the data, then we know the Total variation!

FACTORS (Smoking and Device)
For the SS of the two factors (Smoking SS and Device SS), we treat each factor as if it were the only factor in a One-factor ANOVA, just like we did in the randomized block model. [Note: this only works when we have a *balanced* design, i.e. the sample size in all of the cells is the same.] We take the squared deviation of each level mean from the grand mean, multiply by the appropriate sample size, and sum. Here are the calculations:

Smoking SS
$6(14.933-12.906)^2 = 6 \times 4.1087 = 24.652$
$6(12.883-12.906)^2 = 6 \times 0.0005 = 0.003$
$6(10.900-12.906)^2 = 6 \times 1.0240 = 24.144$
Smoking SS = 48.799

Device SS
$9(10.744-12.906)^2 = 9 \times 4.6742 = 42.068$
$9(15.067-12.906)^2 = 9 \times 4.6699 = 42.029$
Device SS = 84.097

The small differences between the above values and those seen in the ANOVA table on the previous page are due to rounding error (the table was done by computer).

INTERACTION
To understand how interaction variability is calculated, you need to be aware of the following important concept. The idea is that all of the differences among the cell means are *explained*. That is, the reason the cell means differ is due to the two factors and to the interaction. Let's use our example to try to help us understand this. Look at the cell means on the previous page (you'll find them under the data and above the ANOVA table). Why aren't those six numbers exactly the same (i.e. why aren't they all equal to the grand mean of 12.906)? Well, it's because the cell mean is affected by: (1) how much the individuals smoke (the level of the smoking factor); (2) whether they were on the bike or the treadmill (the level of the device factor); (3) and the interaction between smoking and device.

Again, all variability among the cell means is explained - it's due to the factors and their interaction. There is (of course) unexplained variation in this ANOVA model, but it's the *within cell* variability that we can't explain. That's the Error variation that we'll discuss below. Let's get back to interaction.

The way we calculate interaction is to first, calculate how much each cell mean differs from the grand mean. For example, the Bicycle, Nonsmoking mean of 12.5 differs from the grand mean of 12.906 by 12.5 - 12.906 = -0.406 minutes. So, the amount of variability we have to explain is -0.406 minutes.

Second, we calculate the effect of the factors. The effect of using the bicycle is calculated as the difference between the bicycle mean and the grand mean: 10.744 - 12.906 = -2.162. That is, being on the bike caused the subjects to reach maximum O_2 consumption 2.162 minutes before the overall average (grand mean). The effect of being in the Nonsmoking group is the difference between the nonsmoking mean and the grand mean: 14.933 - 12.906 = 2.027. That is, the nonsmokers exercised for 2.027 minutes longer than the overall average (grand mean) before reaching maximum O_2 consumption. To determine the combined effect of using the bike and being a nonsmoker, we simply add the two effects together: -2.162 + 2.027 = -0.135. So, nonsmokers on the bike would be expected to reach maximum O_2 consumption 0.135 minutes before the overall average (grand mean).

Now, combine the results of the two previous paragraphs. The combined effect of the bike and nonsmoking is -0.135 minutes, but the Bicycle, Nonsmoking cell mean differed from the grand mean by -0.406 minutes. How do we explain this discrepancy? It is due to the interaction! The interaction in this cell is the difference between what was observed and what was expected from the combined effect of the factors: -0.406 - (-0.135) = -0.271. So, the interaction between the bike and nonsmoking caused the subjects to reach maximum O_2 consumption 0.271 minutes before the overall average (grand mean).

These calculations are done for each cell. We are now ready to look at the formula for Interaction SS. You should see that this formula incorporates the calculations we performed above:

$$\overline{X} = grand\ mean$$
$$\overline{X}_{ij} = cell\ mean\ (row\ i, column\ j)$$
$$\overline{X}_{i} = level\ mean\ for\ the\ row\ factor\ (e.g.\ device)$$
$$\overline{X}_{j} = level\ mean\ for\ the\ column\ factor\ (e.g.\ smoking\ status)$$
$$n_{ij} = sample\ size\ for\ the\ cells$$
$$a = number\ of\ levels\ in\ the\ row\ factor$$
$$b = number\ of\ levels\ in\ the\ column\ factor$$
$$n_{ij}\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{X}_{ij} - \overline{X} - \left[\left(\overline{X}_{i} - \overline{X}\right) + \left(\overline{X}_{j} - \overline{X}\right)\right]\right)^2$$

Let's apply this formula to our exercise data:

$3\times[(12.500-12.906-[(10.744-12.906)+(14.933-12.906)])^2 +$
$(10.600-12.906-[(10.744-12.906)+(12.883-12.906)])^2 +$
$(9.133-12.906-[(10.744-12.906)+(10.900-12.906)])^2 +$
$(17.367-12.906-[(15.067-12.906)+(14.933-12.906)])^2 +$
$(15.167-12.906-[(15.067-12.906)+(12.883-12.906)])^2 +$
$(12.667-12.906-[(15.067-12.906)+(10.900-12.906)])^2] =$

$3\times[(-0.406 - [-2.162 + 2.027])^2 +$
$(-2.306 - [-2.162 +-0.023])^2 +$
$(-3.773 - [-2.162 +-2.006])^2 +$
$(4.461 - [2.161 + 2.027])^2 +$
$(2.261 - [2.161 +-0.023])^2 +$
$(-0.239 - [2.161 +-2.006])^2 + =$

$3\times[-0.271^2 + -0.121^2 + 0.395^2 + 0.273^2 + 0.123^2 + -0.394^2] =$
$3\times[0.073441 + 0.014641 + 0.156025 + 0.074529 + 0.01529+ 0.155236] =$
$3\times[0.489162] = 1.467 =$ Interaction SS

ERROR
Error is the unexplained variation. What we can't explain in this model is the variation *within* the cells. Within the cells, the subjects used the same device and had the same smoking history, so variation can't be due to the factors. The variability is not explained by our model. The formula and method for calculating Error SS is actually the same idea as in the One-factor ANOVA. What you do is calculate a SS for each cell (i.e. the squared deviation of each datum from its cell mean), and then pool them (i.e. add them together). So **Error SS is a pooled, within cell SS**.

First, we calculate a SS for each cell:

$(12.8-12.500)^2= 0.090$	$(10.9-10.600)^2= 0.090$	$(8.7- 9.133)^2= 0.187$
$(13.5-12.500)^2= 1.000$	$(11.1-10.600)^2= 0.250$	$(9.2- 9.133)^2= 0.004$
$(11.2-12.500)^2= 1.690$	$(9.8-10.600)^2= 0.640$	$(9.5- 9.133)^2= 0.135$
2.780	0.980	0.326

$(17.8-17.367)^2= 0.187$	$(15.5-15.167)^2= 0.111$	$(14.7-12.667)^2= 4.133$
$(18.1-17.367)^2= 0.537$	$(13.8-15.167)^2= 1.869$	$(13.2-12.667)^2= 0.284$
$(16.2-17.367)^2= 1.362$	$(16.2-15.167)^2= 1.067$	$(10.1-12.667)^2= 6.589$
2.086	3.047	11.006

Now we pool (add) the SS values for each cell:
$2.780 + 0.980 + 0.326 + 2.086 + 3.047 + 11.006 = 20.225 =$ Error SS

If you check the ANOVA table, you'll see the computer got 20.2267 as Error SS. Would you care to guess why our calculations above and the computer disagree slightly? Very good! It's rounding error!

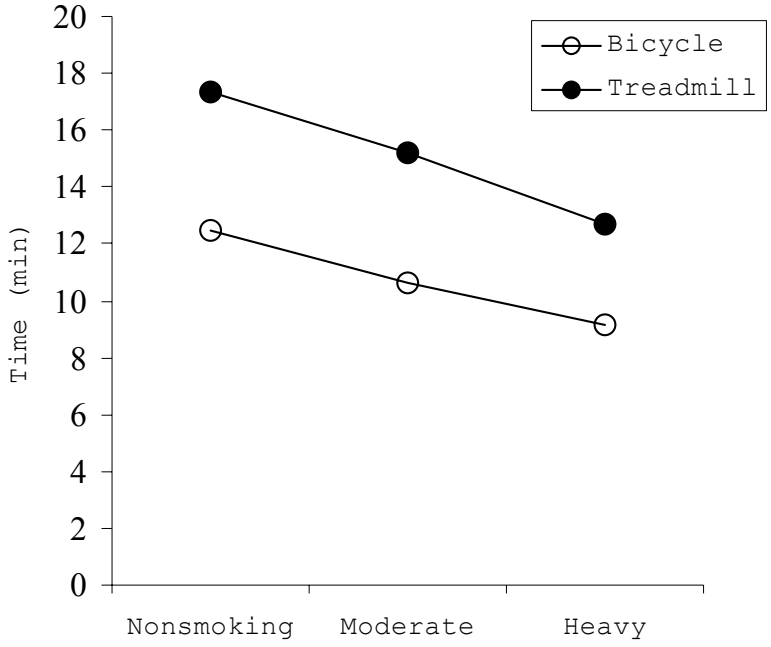
DEGREES OF FREEDOM
Total DF = N - 1 = 18 - 1 = 17. Total DF is N - 1 in all ANOVA models.
Smoking DF = 3 - 1 = 2. Smoking is a factor, so DF = number of levels minus one.
Device DF = 2 - 1 = 1. Device is a factor, so DF = number of levels minus one.
Interaction DF = 2 × 1 = 2. **Interaction DF is the *product* of the degrees of freedom of the factors involved in the interaction.**
Error DF = 17 - 2 - 1 - 2 = 12. Error DF can be calculated by subtraction. Also note that **Error DF is a pooled, within cell DF**. Each cell has 3 data points, so the DF for each cell is 3-1=2. If you pool the six cells: 2+2+2+2+2+2=12.

Now that we have covered all of the sources of variation, let's spend a little time to make sure we understand the concept of interaction. Go to the next page for that discussion.

The null hypothesis of no interaction is accepted, i.e. there is no interaction. This means that as you go from the nonsmoking group, to the moderate smokers, to the heavy smokers, the bike and treadmill subjects responded in the same manner.

As in the randomized block model, it is often easier to visualize this graphically. In the figure below, note that there is a drop in the time to maximum O_2 consumption across the smoking groups on both devices, and that the drop is about the same. In other words, the two lines below are statistically parallel. Thus, there is no significant interaction. A more biological way to say this is: Changes in the time to maximum O_2 consumption for the smoking groups is not dependent on the device used. The amount of change is the same for subjects using the bike and the treadmill.

Note that cell means are graphed here, not individual data points. These cell means have an estimate of error associated with them - the Error term in the ANOVA table.



In a "real-life" research situation, when interactions are present (i.e. you reject the Ho: No Interaction), you will often not want to test for factor effects. For example, if our example was a real study and we had found interaction between smoking groups and the device used, we may not bother testing equality of the smoking group means, or equality of the device means. This is largely due to the fact that our interpretation could be ambiguous, e.g. we couldn't make conclusions about smoking history, because our results across the smoking groups are affected by the device used. We can't untangle the factor effects from the interaction. This is a somewhat simplistic explanation of a complex problem, but hopefully you get the general idea.

On an exam in this class, ***always*** **test for the effects of both factors and for interaction when you have a two-factor ANOVA with replications**. We need the practice in hypothesis testing, and the biological interpretation is not something we have time to deal with in this class.

Experimental Design - Important Safety Tips

If someday you find yourself responsible for designing experimental or observational protocols, you can save time and trouble by planning with statistical testing in mind. Here are a few of the things you should be considering. Since ANOVA is the most widely used statistical test in the biological sciences, many of these remarks specifically relate to that family of analyses.

1. Replicate your cells. If you don't replicate, you cannot measure interaction - but you probably will still have to worry about it.

2. Do a pilot study, or at least write down some numbers like those you expect from your design. Then, actually try to analyze them. What seems like it should work in theory often doesn't when you actually try it.

3. Don't over-factor your design. If you are doing lab work, remember that you should control your experiment. ANOVA can theoretically deal with any number of factors, but in actual practice, you seldom see more than a three-factor ANOVA. This is because:

   A. The sample size needed goes up very rapidly. Let's say you design a 5 factor experiment, where each factor has 4 levels. Since the number of cells is the product of all of the levels, this design has $4\times4\times4\times4\times4 = 1{,}024$ cells. If you decide to have 5 replications in each cell, your sample size is then $1024\times5 = 5{,}120$. If you're working with, say lab mice, you have to deal with 5,120 mice. That's a lot of mouse chow and mouse housing to provide.

   B. The number of hypotheses tested increases rapidly. In our two-factor ANOVA with replications we had three null hypotheses; in a 5-factor ANOVA with replications there would be 31 null hypotheses! This gets too confusing!

4. Try to minimize the number of levels in each factor. As you add levels, you lose statistical power. However, additional levels are often desired because they provide better biological resolution. You may have to compromise.

5. Try to stay with fixed-effects factors (avoid random-effects factors). "Fixed-effects" is where you have measured all the levels that exist, or at least all that are of interest. A "random effect" is one where the levels you have measured represent a subset (a random sample) from a larger population of levels. See Zar for more discussion of fixed and random effects. All of the ANOVA designs we did in class were fixed effects.

6. Avoid unbalanced designs in two-way (or higher) ANOVAs. This is where the sample size within the cells is not the same for every cell. These designs can be analyzed, but the analysis is complicated. You'll need to put in substantial extra time to understand how this is done.

7. Even worse than having unequal sample sizes in the cells are unbalanced designs where some cells have no data points at all. This is extremely complicated, and in many cases, it may result in substantial loss of statistical power.

8. Watch for repeated measures. This is where you measure an object more than once through time, for example, body temperature of an experimental animal is measured every 15 min for two hours. This has to be handled with a special ANOVA design. Don't confuse repeated measures with replications within the cells. Repeated measures are not independent. Repeated measures designs are very common in biology, and if you encounter them in your work you will have to learn to deal with them. Zar has an extensive discussion of these designs in Chapter 14 (which is not part of your assigned reading for the class).

REGRESSION

WHAT IS TESTED
A linear relationship between two variables, where the biological interpretation is that one variable (independent) may biologically cause the other (dependent) variable.

DATA SCALE
Ratio-Interval

NULL HYPOTHESIS
Ho: ß = 0 (No linear relationship)
One-tailed null hypotheses may be tested (see below).

ASSUMPTIONS WITH RESPECT TO DATA
There are several assumptions:

1. Values of the independent variable (X) are fixed and/or measured without error.
2. For each observed value of X, there is a normally distributed population of Y values.
3. The variances of the populations of Y values (see #2 above) are equal.
4. The means of the populations of Y values (see #2 above) lie on a straight line, i.e. the actual relationship in the population is linear.
5. Errors in Y are additive, i.e. for different values of X, the amount that the observed value of Y differs from the "true" value of Y is a linear (not a multiplicative) function of X. The errors are the population residuals; observed Y - true Y. This assumption says that the "true" values of Y can be calculated as follows: $Y_I = a + bX_I + \varepsilon$; where ε is the population residual. Notice you are *adding* the residual to $a + bX_i$, not multiplying.
6. All values of Y are independent of all other values of Y.

DEGREES OF FREEDOM
When testing the null by the analysis of variance technique, the numerator degree of freedom (in BIO 211 class) is always 1; and the denominator degree of freedom is n-2, where n is the number of x,y points.

ZAR CHAPTER
17

TYPE
Parametric

COMMENTS
Regression and correlation are really testing the same null hypothesis in the same way (homologous tests). Although they are mathematically and statistically identical, regression and correlation are often used differently in a biological context. Regression is biologically interpreted as a causal relationship, while correlation is not.

It is critical that you know the important quantities used in the analysis of variance test of the regression null hypothesis, and what they mean. Know TOTAL SS, REGRESSION SS, and RESIDUAL SS as well as their respective MS quantities. Regression problems on exams will be "puzzles", and solving them will require a knowledge of these quantities.

One-tailed null hypotheses (Ho: $\beta \leq 0$ and Ho: $\beta \geq 0$) may only be tested using the *t*-test procedure discussed in Zar (4^{th} edition: pages 336-337; 5^{th} edition: page 341). The ANOVA procedure may not be used for these one-tailed null hypotheses.

Regression - Example

Mom (X) (kg)	Baby (Y) (g)	Predicted Y=1356.4 + 31.0X	Residual (Observed - Predicted)
65.2	3515	3375.6	139.4
58.2	3420	3158.9	261.1
48.7	3175	2864.6	310.4
65.8	3586	3394.2	191.8
73.5	3232	3632.7	-400.7
68.2	3884	3468.6	415.4
69.3	3856	3502.6	353.4
69.3	3941	3502.6	438.4
59.3	3232	3192.9	39.1
73.9	4054	3645.1	408.9
56.3	3459	3100.0	359.0
70.3	3998	3533.6	464.4
62.1	3444	3279.6	164.4
72.1	3827	3589.3	237.7
72.8	3884	3611.0	273.0
49.4	3515	2886.3	628.7
54.4	3416	3041.2	374.8
63.5	3742	3323.0	419.0
61.2	3062	3251.8	-189.8
51.0	3076	2935.9	140.1
44.2	2835	2725.3	109.7
63.1	2750	3310.6	-560.6
63.8	3460	3332.3	127.7
65.8	3340	3394.2	-54.2
59.3	2608	3192.9	-584.9
51.2	2509	2942.1	-433.1
80.0	3600	3834.0	-234.0
60.0	1730	3214.6	-1484.6
74.6	3175	3666.8	-491.8
68.7	3459	3484.0	-25.0
69.7	3288	3515.0	-227.0
62.3	2920	3285.8	-365.8
65.1	3020	3372.5	-352.5
49.9	2778	2901.8	-123.8
46.7	2466	2802.7	-336.7
61.2	3260	3251.8	8.2
Mean 62.5	3292.1	3292.1	0.0
Standard Deviation 8.8	499.9		

ANOVA for Regression Ho: β=0

SOURCE	SS	DF	MS
Total	8,747,373.6	35	249,925.0
Regression	2,623,275.6	1	2,623,275.6
Error (Residual)	6,124,097.9	34	180,120.5

$$F = \frac{\text{Regression MS}}{\text{Error MS}} = \frac{2{,}623{,}275.6}{180{,}120.5} = 14.564$$ DF = 1, 34 p<0.05 (p=.00055) Reject Ho:

Coefficient of Determination = r^2 = proportion of total variability explained

$$r^2 = \frac{\text{Regression SS}}{\text{Total SS}} = \frac{2{,}623{,}275.6}{8{,}747{,}373.6} = 0.3 = 30\%$$

30% of the variation in baby weight is explained by mom's weight
70% of the variation in baby weight is due to other factors

Standard Error of Estimate $= \sqrt{\text{Residual MS}} = \sqrt{180{,}120.5} =$ 424.4 g

$$b = \frac{\sum \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sum \left(X_i - \overline{X}\right)^2}$$

The slope of the regression line (b) is calculated by this formula. The numerator is called the "sum of the crossproducts". The denominator is simply the sum of squares of the independent (X) variable. Here's the calculation of b for the baby data.

$\sum \left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)$	$\sum \left(X_i - \overline{X}\right)^2$
(65.2-62.5)(3515-3292.1)= 2.7 × 222.9 = 602	$(65.2-62.5)^2$= 7
(58.2-62.5)(3420-3292.1)= -4.3 × 127.9 = -550	$(58.2-62.5)^2$= 18
(48.7-62.5)(3175-3292.1)= -13.8 × -117.1 = 1616	$(48.7-62.5)^2$= 190
(65.8-62.5)(3586-3292.1)= 3.3 × 293.9 = 970	$(65.8-62.5)^2$= 11
(73.5-62.5)(3232-3292.1)= 11.0 × -60.1 = -661	$(73.5-62.5)^2$= 121
(68.2-62.5)(3884-3292.1)= 5.7 × 591.9 = 3374	$(68.2-62.5)^2$= 32
(69.3-62.5)(3856-3292.1)= 6.8 × 563.9 = 3835	$(69.3-62.5)^2$= 46
(69.3-62.5)(3941-3292.1)= 6.8 × 648.9 = 4413	$(69.3-62.5)^2$= 46
(59.3-62.5)(3232-3292.1)= -3.2 × -60.1 = 192	$(59.3-62.5)^2$= 10
(73.9-62.5)(4054-3292.1)= 11.4 × 761.9 = 8686	$(73.9-62.5)^2$= 130
(56.3-62.5)(3459-3292.1)= -6.2 × 166.9 = -1035	$(56.3-62.5)^2$= 38
(70.3-62.5)(3998-3292.1)= 7.8 × 705.9 = 5506	$(70.3-62.5)^2$= 61
(62.1-62.5)(3444-3292.1)= -0.4 × 151.9 = -61	$(62.1-62.5)^2$= 0
(72.1-62.5)(3827-3292.1)= 9.6 × 534.9 = 5135	$(72.1-62.5)^2$= 92
(72.8-62.5)(3884-3292.1)= 10.3 × 591.9 = 6097	$(72.8-62.5)^2$= 106
(49.4-62.5)(3515-3292.1)= -13.1 × 222.9 = -2920	$(49.4-62.5)^2$= 172
(54.4-62.5)(3416-3292.1)= -8.1 × 123.9 = -1004	$(54.4-62.5)^2$= 66
(63.5-62.5)(3742-3292.1)= 1.0 × 449.9 = 450	$(63.5-62.5)^2$= 1
(61.2-62.5)(3062-3292.1)= -1.3 × -230.1 = 299	$(61.2-62.5)^2$= 2
(51.0-62.5)(3076-3292.1)= -11.5 × -216.1 = 2485	$(51.0-62.5)^2$= 132
(44.2-62.5)(2835-3292.1)= -18.3 × -457.1 = 8365	$(44.2-62.5)^2$= 335
(63.1-62.5)(2750-3292.1)= 0.6 × -542.1 = -325	$(63.1-62.5)^2$= 0
(63.8-62.5)(3460-3292.1)= 1.3 × 167.9 = 218	$(63.8-62.5)^2$= 2
(65.8-62.5)(3340-3292.1)= 3.3 × 47.9 = 158	$(65.8-62.5)^2$= 11
(59.3-62.5)(2608-3292.1)= -3.2 × -684.1 = 2189	$(59.3-62.5)^2$= 10
(51.2-62.5)(2509-3292.1)= -11.3 × -783.1 = 8849	$(51.2-62.5)^2$= 128
(80.0-62.5)(3600-3292.1)= 17.5 × 307.9 = 5388	$(80.0-62.5)^2$= 306
(60.0-62.5)(1730-3292.1)= -2.5 ×-1562.1 = 3905	$(60.0-62.5)^2$= 6
(74.6-62.5)(3175-3292.1)= 12.1 × -117.1 = -1417	$(74.6-62.5)^2$= 146
(68.7-62.5)(3459-3292.1)= 6.2 × 166.9 = 1035	$(68.7-62.5)^2$= 38
(69.7-62.5)(3288-3292.1)= 7.2 × -4.1 = -30	$(69.7-62.5)^2$= 52
(62.3-62.5)(2920-3292.1)= -0.2 × -372.1 = 74	$(62.3-62.5)^2$= 0
(65.1-62.5)(3020-3292.1)= 2.6 × -272.1 = -707	$(65.1-62.5)^2$= 7
(49.9-62.5)(2778-3292.1)= -12.6 × -514.1 = 6478	$(49.9-62.5)^2$= 159
(46.7-62.5)(2466-3292.1)= -15.8 × -826.1 = 13052	$(46.7-62.5)^2$= 250
(61.2-62.5)(3260-3292.1)= -1.3 × -32.1 = 42	$(61.2-62.5)^2$= 2
-------	-----
Σ = 84703	Σ = 2735

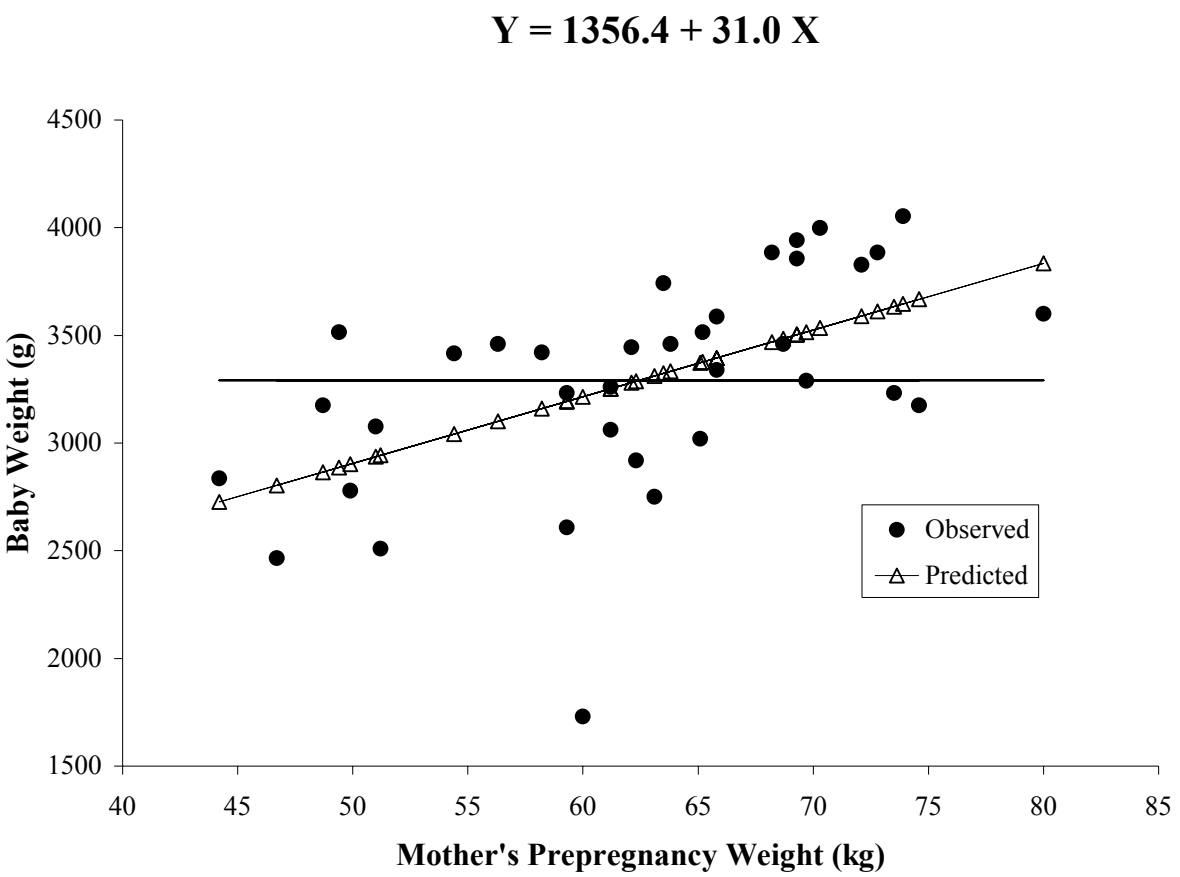
$$b = \frac{84703}{2735} = 31.0$$

This is the value for b that satisfies the **least squares linear regression criteria,** i.e. it minimizes Residual SS. (The calculation of Residual SS will be seen in a couple of pages.) Now that we know b, we can calculate the intercept (a) by this method:

$$\overline{Y} = a + b\overline{X},\ \therefore a = \overline{Y} - b\overline{X}, \quad a = 3292.1 - \left(31.0 \times 62.5\right) = 1356.4$$

If you do the above calculation for the intercept (a) on your calculator, you'll get 1354.6. The difference is due to (you guessed it) rounding error.

Regression - Example - Scatter Plot



The horizontal line in the above figure is at Y = 3292. Remember that 3292 is the mean of the Y values, i.e. the mean baby weight.

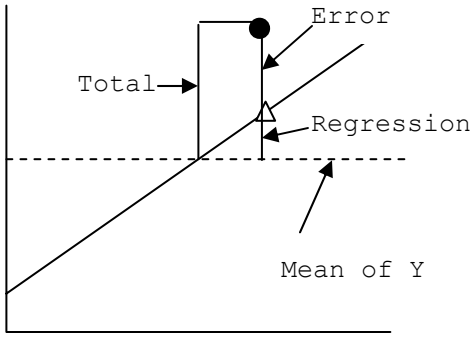
TOTAL SS is the sum of the squared distances between each observed weight (black circles) and the mean of Y (3292).

REGRESSION SS is the sum of the squared distances between each predicted weight (open triangles on the regression line) and the mean of Y (3292).

RESIDUAL (ERROR) SS is the sum of the squared distances between each observed weight (black circle) and the associated predicted weight (open triangle).

See the next page for the calculations.

The figure on the right is a graphical representation (not mathematically accurate) of partitioning in regression. Total is observed variability; regression is predicted; and error (residual) is the difference between observed and predicted. Total SS = Regression SS + Error SS.



Sources of Variation - ANOVA for Regression - Example

Total SS = $\sum_{i=1}^{n}(Y_i-\overline{Y})^2$	*Regression SS* = $\sum_{i=1}^{n}(\hat{Y}-\overline{Y})^2$	*Residual SS* = $\sum_{i=1}^{n}(Y_i-\hat{Y})^2$
$(3515-3292)^2 = 49729$	$(3375.6-3292)^2 = 6996.5$	$(3515-3375.6)^2 = 19419.8$
$(3420-3292)^2 = 16384$	$(3158.9-3292)^2 = 17728.2$	$(3420-3158.9)^2 = 68197.9$
$(3175-3292)^2 = 13689$	$(2864.6-3292)^2 = 182641.1$	$(3175-2864.6)^2 = 96326.6$
$(3586-3292)^2 = 86436$	$(3394.2-3292)^2 = 10450.4$	$(3586-3394.2)^2 = 36776.8$
$(3232-3292)^2 = 3600$	$(3632.7-3292)^2 = 116075.5$	$(3232-3632.7)^2 = 160559.4$
$(3884-3292)^2 = 350464$	$(3468.6-3292)^2 = 31172.0$	$(3884-3468.6)^2 = 172593.8$
$(3856-3292)^2 = 318096$	$(3502.6-3292)^2 = 44362.2$	$(3856-3502.6)^2 = 124875.1$
$(3941-3292)^2 = 421201$	$(3502.6-3292)^2 = 44362.2$	$(3941-3502.6)^2 = 192174.2$
$(3232-3292)^2 = 3600$	$(3192.9-3292)^2 = 9816.8$	$(3232-3192.9)^2 = 1527.2$
$(4054-3292)^2 = 580644$	$(3645.1-3292)^2 = 124670.2$	$(4054-3645.1)^2 = 167210.1$
$(3459-3292)^2 = 27889$	$(3100.0-3292)^2 = 36860.5$	$(3459-3100.0)^2 = 128874.5$
$(3998-3292)^2 = 498436$	$(3533.6-3292)^2 = 58367.5$	$(3998-3533.6)^2 = 215673.3$
$(3444-3292)^2 = 23104$	$(3279.6-3292)^2 = 152.8$	$(3444-3279.6)^2 = 27015.2$
$(3827-3292)^2 = 286225$	$(3589.3-3292)^2 = 88411.2$	$(3827-3589.3)^2 = 56482.2$
$(3884-3292)^2 = 350464$	$(3611.0-3292)^2 = 101773.4$	$(3884-3611.0)^2 = 74518.4$
$(3515-3292)^2 = 49729$	$(2886.3-3292)^2 = 164581.2$	$(3515-2886.3)^2 = 395246.2$
$(3416-3292)^2 = 15376$	$(3041.2-3292)^2 = 62918.0$	$(3416-3041.2)^2 = 140500.9$
$(3742-3292)^2 = 202500$	$(3323.0-3292)^2 = 960.7$	$(3742-3323.0)^2 = 175564.9$
$(3062-3292)^2 = 52900$	$(3251.8-3292)^2 = 1619.0$	$(3062-3251.8)^2 = 36010.2$
$(3076-3292)^2 = 46656$	$(2935.9-3292)^2 = 126831.1$	$(3076-2935.9)^2 = 19637.4$
$(2835-3292)^2 = 208849$	$(2725.3-3292)^2 = 321184.9$	$(2835-2725.3)^2 = 12041.1$
$(2750-3292)^2 = 293764$	$(3310.6-3292)^2 = 346.2$	$(2750-3310.6)^2 = 314280.5$
$(3460-3292)^2 = 28224$	$(3332.3-3292)^2 = 1623.0$	$(3460-3332.3)^2 = 16310.7$
$(3340-3292)^2 = 2304$	$(3394.2-3292)^2 = 10450.4$	$(3340-3394.2)^2 = 2940.6$
$(2608-3292)^2 = 467856$	$(3192.9-3292)^2 = 9816.8$	$(2608-3192.9)^2 = 342131.5$
$(2509-3292)^2 = 613089$	$(2942.1-3292)^2 = 122457.7$	$(2509-2942.1)^2 = 187541.4$
$(3600-3292)^2 = 94864$	$(3834.0-3292)^2 = 293770.2$	$(3600-3834.0)^2 = 54758.7$
$(1730-3292)^2 = 2439844$	$(3214.6-3292)^2 = 5990.9$	$(1730-3214.6)^2 = 2204035.0$
$(3175-3292)^2 = 13689$	$(3666.8-3292)^2 = 140449.5$	$(3175-3666.8)^2 = 241833.8$
$(3459-3292)^2 = 27889$	$(3484.0-3292)^2 = 36879.8$	$(3459-3484.0)^2 = 627.1$
$(3288-3292)^2 = 16$	$(3515.0-3292)^2 = 49734.1$	$(3288-3515.0)^2 = 51534.2$
$(2920-3292)^2 = 138384$	$(3285.8-3292)^2 = 38.1$	$(2920-3285.8)^2 = 133832.3$
$(3020-3292)^2 = 73984$	$(3372.5-3292)^2 = 6488.0$	$(3020-3372.5)^2 = 124290.0$
$(2778-3292)^2 = 264196$	$(2901.8-3292)^2 = 152256.8$	$(2778-2901.8)^2 = 15326.2$
$(2466-3292)^2 = 682276$	$(2802.7-3292)^2 = 239420.3$	$(2466-2802.7)^2 = 113362.9$
$(3260-3292)^2 = 1024$	$(3251.8-3292)^2 = 1619.0$	$(3260-3251.8)^2 = 67.8$
Total SS = 8747374	Regression SS = 2623276.1	Residual SS = 6124097.9

If you compare the above SS quantities to those you saw earlier in the ANOVA table, you may notice very small differences. Yes, this is due to rounding error. The ANOVA table was generated by a computer.

DEGREES OF FREEDOM
The relevant sample size is n = the number of X,Y data points. n = 36 for our example. Note: there are 36 moms (X) and 36 babies (Y), and that n = 36 (not 72!)

Total DF = n- 1. For our example, Total DF = 36 - 1 = 35.
Regression DF = 1. This is true for all the regressions we do in BIO 211, but not for all regressions everywhere.
Residual (Error) DF = n - 2. For our example, Residual (Error) DF = 36 - 2 = 34.

Note that, as always, DF values are additive:
Residual DF + Regression DF = Total DF (For our example: 34 + 1 = 35)

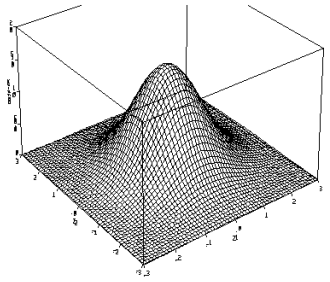
CORRELATION

WHAT IS TESTED
Linear relationship (no causal relationship assumed) between two variables.

DATA SCALE
Ratio-Interval

NULL HYPOTHESIS
Ho: ρ=0 (ρ is the Greek letter Rho in lower case)
One-tailed null hypotheses may be tested.

ASSUMPTIONS WITH RESPECT TO DATA
The x,y points are assumed to be a random sample from a normally distributed population of x,y points (bivariate normal distribution).
Here is a graphical representation of a bivariate normal distribution.



DEGREES OF FREEDOM
DF = n-2 where n is the number of x,y points

ZAR CHAPTER
19

TYPE
Parametric. Nonparametric analogue is the Spearman Rank Correlation procedure.

COMMENTS
Correlation is really the same test as Regression. The correlation coefficient (r) is actually the regression coefficient (slope of the line, i.e. b) between two standardized (z-scores) variables. The difference between correlation and regression is in the biological interpretation (no causal relationship in correlation).

The Pearson product-moment correlation coefficient is calculated by this formula:

$$r = \frac{\sum\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum\left(X_i - \overline{X}\right)^2 \sum\left(Y_i - \overline{Y}\right)^2}} = \frac{84703}{\sqrt{2735 \times 8747374}} = 0.548$$

Notice that we calculated all of these quantities when we did Regression analysis. The numerator is the sum of crossproducts, the denominator is the square root of the product of the sum of squares of the X variable times the sum of squares of the Y variable. Remember that $-1 \leq r \leq 1$, and that the closer the $|r|$ is to 1, the more linear the data. An $|r|=1$ is a perfect straight line.

As you can see above, the correlation between the moms' prepregnancy weights and the babys' weights is r = 0.548. Check the significance of this coefficient in Zar, Table B.17 (4^{th} edition: page App109; 5^{th} edition: page 766). DF = n-2 = 36-2 = 34; $p<0.05$, so we reject Ho: ρ=0.

In fact, an r of 0.548 with 34 DF has an $\alpha(2)$ probability of 0.00055. This is exactly the same alpha probability we had when we tested the Ho: β=0 in the regression on the same data. Why is the alpha probability in regression and in correlation the same? Because they're the same test!

SPEARMAN RANK CORRELATION

WHAT IS TESTED
Linear relationship between two variables.

DATA SCALE
Ordinal

NULL HYPOTHESIS
Ho: $\rho_s=0$ (ρ is the Greek letter Rho in lower case)
One-tailed null hypotheses may be tested.

ASSUMPTIONS WITH RESPECT TO DATA
None

DEGREES OF FREEDOM
When looking in the table, use n, the number of x,y data points. There is no degree of freedom calculation.

ZAR CHAPTER
19

TYPE
Nonparametric. Parametic analogue is Correlation.

COMMENTS
The Spearman Rank Correlation is about 91% as powerful as the Correlation (Pearson product-moment) procedure.

Rank the data of the two variables independently. Be careful not to become confused with the ranking method used in the Mann-Whitney U test.

When calculating the Spearman rank correlation coefficient, be very careful. It is easy to make mistakes evaluating the formula. Be sure to practice this before exams.

Watch your subscripts! r_s and ρ_s refer to the Spearman quantity, while r and ρ refer to the Pearson product-moment quantity. They are not the same! Be careful!

Spearman Rank Correlation – Example

For this example, we will use the same Mom and Baby data as in regression and correlation. **Null Hypothesis Ho: $\rho_s = 0$**

Mom	Rank	Baby	Rank	d_i	d_i^2
65.2	22	3515	25.5	-3.5	12.25
58.2	10	3420	20	-10	100
48.7	3	3175	12.5	-9.5	90.25
65.8	23.5	3586	27	-3.5	12.25
73.5	33	3232	14.5	18.5	342.25
68.2	25	3884	32.5	-7.5	56.25
69.3	27.5	3856	31	-3.5	12.25
69.3	27.5	3941	34	-6.5	42.25
59.3	11.5	3232	14.5	-3	9
73.9	34	4054	36	-2	4
56.3	9	3459	22.5	-13.5	182.25
70.3	30	3998	35	-5	25
62.1	16	3444	21	-5	25
72.1	31	3827	30	1	1
72.8	32	3884	32.5	-0.5	0.25
49.4	4	3515	25.5	-21.5	462.25
54.4	8	3416	19	-11	121
63.5	19	3742	29	-10	100
61.2	14.5	3062	10	4.5	20.25
51.0	6	3076	11	-5	25
44.2	1	2835	7	-6	36
63.1	18	2750	5	13	169
63.8	20	3460	24	-4	16
65.8	23.5	3340	18	5.5	30.25
59.3	11.5	2608	4	7.5	56.25
51.2	7	2509	3	4	16
80.0	36	3600	28	8	64
60.0	13	1730	1	12	144
74.6	35	3175	12.5	22.5	506.25
68.7	26	3459	22.5	3.5	12.25
69.7	29	3288	17	12	144
62.3	17	2920	8	9	81
65.1	21	3020	9	12	144
49.9	5	2778	6	-1	1
46.7	2	2466	2	0	0
61.2	14.5	3260	16	-1.5	2.25

NOTE: Be sure to rank the two variables independently. In this example, rank the Moms by themselves; and then the Babys by themselves. You can rank "low to high" or "high to low" – just do both variables the same way.

NOTE: d_i is the difference between the ranks – not between the original data. The order in which you do the subtraction (i.e. the sign on d_i) does not matter, because you square the d_i value.

$\sum d_i^2 = 3065$

$$r_s = 1 - \frac{6 \times \sum_{i=1}^{n} d_i^2}{n^3 - n} = 1 - \frac{6 \times 3065}{36^3 - 36} = 1 - \frac{18390}{46656 - 36} = 1 - \frac{18390}{46620} = 1 - 0.3945 = 0.6055$$

The critical value at $\alpha(2)=0.05$ for n=36 is 0.330 (Zar Table B.20. 4th edition: page App116; 5th edition: page 773)

Our calculated value exceeds the critical value, therefore the alpha probability associated with our calculated value of 0.6055 is less than 0.05. ($p < 0.05$)

Therefore, we reject the null hypothesis. There is a significant relationship between baby's birth weight and mother's pre-pregnancy weight.

Be careful when you use this formula. It's easy to make mistakes.

ANALYSIS OF COVARIANCE (ANCOVA)

WHAT IS TESTED
Difference among two or more adjusted means. The means are adjusted by using the linear relationship (i.e. regression) with a second variable (the *covariate*).

DATA SCALE
Ratio - Interval

NULL HYPOTHESIS
Ho: Equality of two or more adjusted means

ASSUMPTIONS WITH RESPECT TO DATA
Related to the assumptions of ANOVA and Regression. Do not be concerned with the details here - time will not permit detailed analysis.

DEGREES OF FREEDOM
Depends on the details of the particular analysis. Again, do not be concerned with this detail.

ZAR CHAPTER
There is a very brief introduction in Zar Chapter 12 (4^{th} edition: pages 270-271; 5^{th} edition: page 284). However, be sure to read the following pages in the Test Pac. These pages contain what you need to know.

TYPE
Parametric

COMMENTS
This is a complex analysis that we will examine only superficially. Read the following example in the Test Pac. Try to see what is being tested and to follow what is happening in a general sense. Do not be concerned with the mathematical or statistical details here. Try to grasp the concept of ANCOVA.

You will not be asked to do an ANCOVA problem on the open book portion of an exam. In fact, you will not be given sufficient information to do an ANCOVA problem, even if you wanted to. However, you should expect to be asked questions about ANCOVA on the closed book (multiple choice) portion of an exam.

ANCOVA (Analysis of Covariance) - Example

In our classroom examples over the past few weeks, we have examined aspects of infant birth weights. Namely:

1. With one-factor ANOVA, we found a significant difference in birth weight among babies grouped by smoking habits of the mother during pregnancy.

2. With regression, we found that birth weight is a linear function of mother's pre-pregnancy body weight. In fact, about 30% of the variance in birth weight is due to body weight of moms.

Now, consider the possible effect of our regression results on the interpretation of our ANOVA results. When we did the ANOVA, we did not control for the body weight of the mother. What if the 1+ Pack/Day moms just happened to be small? Then, the low birth weight of their infants might be due to their body size and not to their smoking habits. In other words, mom's bodyweight is a **confounding variable** (or **concomitant variable)** in our investigation of smoking habits. What we need to do is account for the effect of mom's body weight.

Analysis of Covariance (ANCOVA) is designed to do what we need. ANCOVA will do two things for us: (1) adjust the means of the birth weights in the smoking groups (nonsmokers; 1 pack/day; 1+ pack/day) using the body weights of the moms. This adjustment is done using the regression relationship. It has the effect of holding the mom's weight constant, i.e. it statistically answers the question "What would the birth weights have been if all 36 moms had the same pre-pregnancy weight?"; (2) adjust the ERROR SS in the ANOVA. Remember that ERROR SS (within smoking group variability) is "unexplained" variance. But now we can explain some of the within group variability - 30% of it can be explained by regression on mom's weight.

ANCOVA combines the techniques of ANOVA and regression. Here is a summary of what the ANCOVA does.

The first step is to see if there is a linear relationship between baby's weight and mom's weight for each of the three smoking groups. On the next page is a plot of the data with regression lines for each group.

For all three regressions, Ho: ß = 0 is rejected ($p<.05$).

The plot on the next page shows baby's birth weight as a function of mother's pre-pregnancy weight for each of the three smoking groups.

The data points represented by the filled circles (●)on the plot are the nonsmoking moms and babies. The regression equation for nonsmokers is:

$$Y = 1812.5 + 27.8 X$$

The data points represented by the open circles (○) on the plot are the 1 pack/day moms and babies. The regression equation for the 1 pack/day group is:

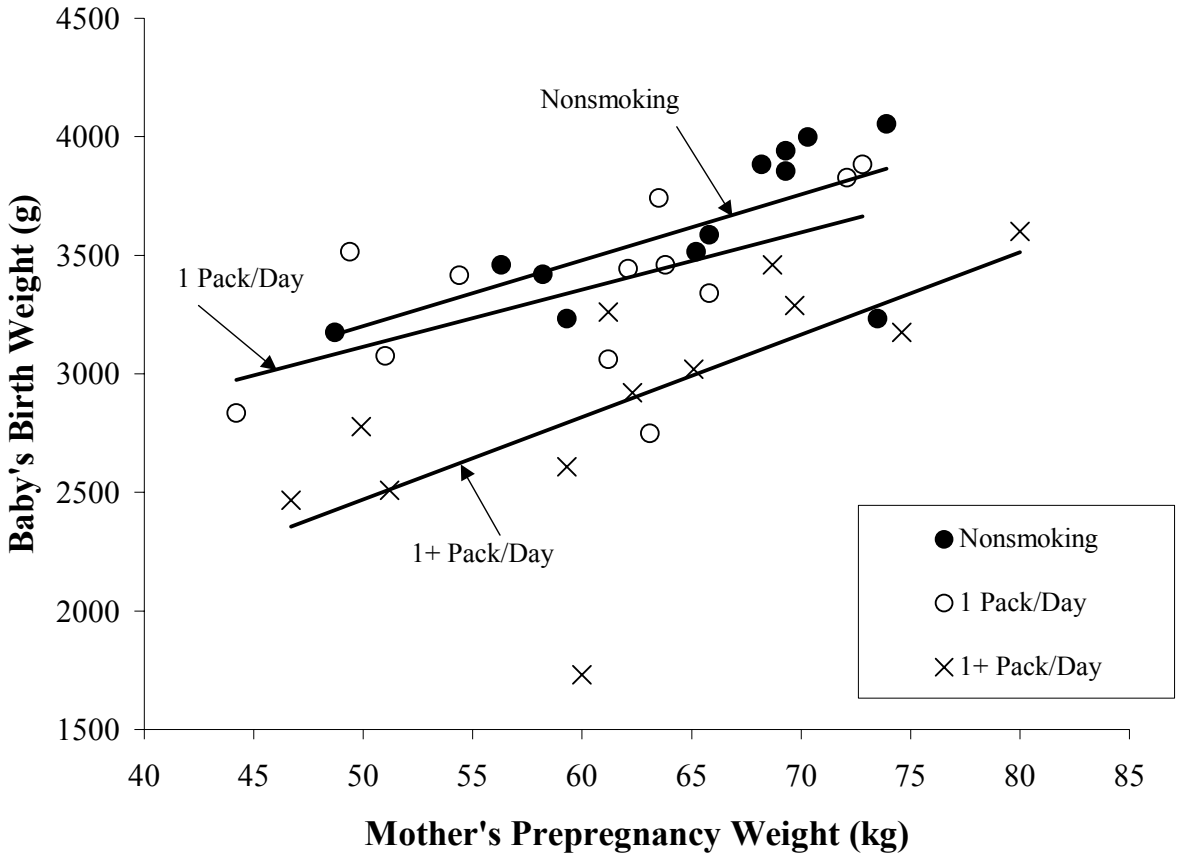
$$Y = 1905.6 + 24.2 X$$

The data points represented by the letter X on the plot are the 1+ pack/day moms and babies. The regression equation for 1+ pack/day group is:

$$Y = 733.5 + 34.7 X$$

The lines generated by these regression equations are indicated with arrows.

ANCOVA - Example plot



In order for the ANCOVA to proceed, the slopes of the three regression lines must not be statistically different. The slopes appear very similar, and next we need to test this assumption. The test for equality of slopes is in Zar, but not in a chapter you were assigned to read (it's Chapter 18). In this case the null hypothesis of equal slopes (Ho: $\beta_1 = \beta_2 = \beta_3$) is accepted ($p > 0.05$, in fact the alpha probability in the test was p=0.77).

The ANCOVA procedure now calculates an "adjusted mean" for the baby weights in the smoking groups. The means are adjusted based on: (1) the linear relationship between birth weight and mom's weight; and (2) how far the mean of the group of moms is from the mean of all moms.

The actual equation used to adjust the means is: $\overline{B}_{adj} = \overline{B}_i - b_p\left(\overline{M}_i - \overline{M}\right)$

In this equation, the Bs refer to the babies; Ms refer to the moms, and b_p is the regression coefficient (called the "pooled" regression coefficient in this analysis). The M without the subscript refers to grand mean of all moms.

For example, the adjusted mean for the nonsmoking babies is calculated as:

Adjusted Mean = 3613 - 29.53(65 - 63) = 3544

Examine the table that follows, which shows the original mean birth weights, moms' mean weights, and adjusted mean birth weights.

Group	Original Mean of Baby Weights (g)	Mean Weights of Moms (kg)	Adjusted Mean of Baby Weights (g)
Nonsmokers (n=12)	3613	65	3544
1 Pack/day (n=12)	3363	60	3428
1+ Pack/day (n=12)	2901	62	2904
Grand Mean (n=36)	3292	63	3292

Notice that the nonsmoking moms were 2 kg heavier than all the moms. Therefore, their babies may average a little heavier just due to mom's size. Regression tells us that this mean should be adjusted down about 70 g to account for the larger moms. The 1 pack/day moms were 3 kg lighter than all moms, therefore, the mean weight of their babies is adjusted up about 65 g. Similarly, the 1+ pack/day group is adjusted up slightly. ANCOVA tests for a significant difference between these adjusted means.

Just for the purposes of accuracy and completeness, here are the calculations of the adjusted means, using four decimal places:

```
  Nonsmoking:     3612.6667 - 29.5317(64.8333 - 62.5028) = 3543.8422
  1 Pack/Day:     3362.5833 – 29.5317(60.2833 – 62.5028) = 3428.1280
 1+ Pack/Day:     2901.0833 – 29.5317(62.3917 – 62.5028) = 2904.3643
```

When ANCOVA tests the adjusted means, it "partitions out" the variance in the birth weight data which can be explained by regression on moms' weight. That quantity is what we call Regression SS (Regression MS) in regression analysis, i.e. the variation in birth weight (Y) explained by the linear relationship with mom's pre-pregnancy weight (X). Here is the ANCOVA table:

Source	SS	DF	MS	F	p
Total (pooled)	8,400,802.0	35			
Adjusted means	2,780,927.5	2	1,390,463.8	13.309	0.00006
Regression (pooled)	2,276,706.5	1	2,276,706.5	21.792	0.00005
Error	3,343,168.0	32	104,474.0		

The F = 13.309 tests for equality of the adjusted means, and clearly the hypothesis is rejected. That is, there is a significant difference between the adjusted means. When we tested the original means with one-factor ANOVA, our F value was 9.18 (with 2 and 33 degrees of freedom), which has a p = 0.0007. Although we rejected the null in both analyses, notice that our alpha probability with the ANCOVA is an order of magnitude smaller. ANCOVA has given us a more powerful test of the differences among the smoking groups, and we are also now sure that our results are not due to differences among the weights of the moms.

The F = 21.792 tests for a significant regression of baby weight on moms' weights over all data (i.e. all three groups combined, n=36). The "pooled regression" approach holds the smoking factor constant. Therefore, there is a highly significant linear relationship between birth weight and moms weights, holding constant mom's smoking status.

NOTE: Notice in the ANCOVA table above that several sources are labeled "pooled". This refers to the ANCOVA method, the details of which will not be covered.

The following page compares ANOVA, Regression, and ANCOVA.

The pictures below show the partitioning of Total SS in the ANOVA, Regression, and ANCOVA. Remember that the same variable (birth weight of the 36 babies) was the dependent variable in each analysis.

In ANOVA, we tried to explain variability in birth weight using smoking status of mom during pregnancy.

In Regression, we tried to explain variability in birth weight using mom's pre-pregnancy body weight.

In ANCOVA, we tried to explain variability in birth weight using smoking status of mom during pregnancy AND mom's pre-pregnancy body weight.

ANOVA

Smoking (Groups) SS 3,127,499	Error SS 5,619,874

Regression

Regression SS 2,623,276	Error (Residual) SS 6,124,098

ANCOVA

Smoking SS (Adjusted Means) 2,780,928	Regression SS 2,276,707	Error SS 3,343,168

Notice that Smoking SS was higher in the ANOVA than in the ANCOVA. We explained more birth weight variability with Smoking status of mom in the ANOVA than we did in the ANCOVA. This makes sense if you look at the table on the previous page. The adjusted means are closer together than the original means. Since Smoking SS is a Groups SS, it measures how far apart the means of the groups are. So it should be higher in the ANOVA.

Notice that Regression SS was higher in the Regression than in the ANCOVA. We explained more birth weight variability with mom's pre-pregnancy weight in the Regression than we did in the ANCOVA. This makes sense, because Regression SS is a measure of the magnitude of the slope. In the Regression, the slope of the line was 31.0, but in the ANCOVA, the pooled slope was 29.5. Since the slope was larger in the Regression, the Regression SS should be higher.

If Smoking SS was higher in the ANOVA; and Regression SS was higher in the Regression, then what's the advantage of the ANCOVA? From the above diagrams showing partitioning, it is obvious that Error SS is MUCH smaller in the ANCOVA. Although Smoking SS and Regression SS were smaller in the ANCOVA, the fact that the ANCOVA uses both independent variables results in much more variability being explained (Smoking SS + Regression SS = 2,780,928 + 2,276,707 = 5,057,635) than either the ANOVA or Regression. The combination of the two predictors is much better than either is alone – this should make sense – two predictors are better than one. Not only do we get a better prediction, but remember that ANCOVA shows us the effect of each predictor holding the other one constant.

ANCOVA is a very important analysis in biological research.

THE CENTRAL LIMIT THEOREM - INTRODUCTION

The central limit theorem is a very important theorem in mathematics, statistics, and all areas of science. The proof and the mathematical details of the theorem are best left to the mathematicians. We will concentrate on what is important for biologists. We will examine what the central limit theorem tells us about the: (1) distribution of complex variables; and (2) sampling distribution of the mean.

The **Distribution of Complex Variables**

The central limit theorem says that any variable which is affected by many other, independent variables will tend to be normally distributed. In other words, complex variables usually are expected to have a normal distribution. Since most biological variables are complex, most will tend to a normal distribution.

Consider, for example, heart beat rate in humans. The rate at which a person's heart beats is affected by a number of factors including: age, sex, size, physical condition, emotional condition, past medical history, and random genetic factors. In other words, heart rate is the product of a number of independent variables. The central limit theorem predicts that heart rate should be normally distributed.

The applicability of this aspect of the theorem is so general that it is common to see demonstrations of it in museums of science and industry. The demonstrations (called Galton boards after their inventor, Sir Francis Galton) usually involve dropping colored balls from a funnel. As the balls fall, they hit a series of pegs and bounce all over, eventually falling into a series of slots. The distribution of the balls in the slots (i.e. number of balls/slot) tends to be normal, with the mean of the distribution under the exit point of the funnel. The pegs represent a large number of independent variables which affect the particular slot into which a ball will fall.

The **Sampling Distribution of the Mean**

The central limit theorem provides us with critical information about the sampling distribution of the mean. Before discussing this information, let's be sure we understand what the sampling distribution of the mean is. Remember that the term "sampling distribution" refers to a particular type of frequency distribution, i.e. a frequency distribution of the values of a statistic. The sampling distribution of the mean is therefore a frequency distribution showing how often different values of the mean occur. Suppose we took 100 samples all of the same size from some population, and we calculated 100 means. If we constructed a table or graph showing how often various values of the mean occurred, this would be a sampling distribution of the mean.

The central limit theorem tells us the following three things about the sampling distribution of the mean:

1. If we take repeated samples from a population that is normally distributed, and construct a sampling distribution of the means of those samples, the sampling distribution will be normally distributed.

2. If we take samples from a nonnormal population, the sampling distribution of the mean will approach a normal distribution. It will get closer for larger sample sizes (i.e. it is closer to normal when we take repeated samples of size n=20 than when we take samples of size n=5).

3. If we are taking repeated samples from a population, the standard deviation of the sampling distribution (i.e. the standard error) will decrease as the sample size increases. In other words, when we use large sample sizes, our estimates of the mean are more accurate and therefore more similar to one another (i.e. less variable).

The central limit theorem also tells us that the best estimate of the standard error (i.e. the standard deviation of the sampling distribution) of the mean (when we have taken just one sample) is to take the standard deviation of the sample and divide it by the square root of the sample size.

The above information about the sampling distribution of the mean is usually not easily understood. Don't worry if you are confused after you read this the first time. Go through the material several times, and it will also be discussed in lecture. To further help you, let's consider an example. The illustration that follows this section is for the example; refer to the illustration as it is discussed.

Suppose we have two populations: (1) a normally distributed population with a mean (μ) of 50.5 and a standard deviation (σ) of 10; and (2) a uniform distribution of all the integers between 1 and 100. "Uniform" means each of the integers occurs with equal frequency. The mean (μ) of this uniform distribution is also 50.5, and the standard deviation (σ) is 29. At the top of the illustration we see our populations. The uniform distribution appears as a straight line, i.e. all values of x have equal frequency. (The y-axis is frequency; x-axis is the data value.)

Remember that the above two distributions are data distributions, they are not sampling distributions because they don't show how often values of statistics occur. They show how often values of data occur.

Now we are going to take samples from these two populations. We have the computer take 5000 samples of size n=2; 5000 samples of size n=5; and 5000 samples of size n=30 from **each** population. The computer calculates the mean of each sample; remember that these sample means are estimates of the population mean, which we know is 50.5 for both of our populations.

The distributions underneath the populations are **sampling distributions of the mean**. That is, they show us how often different values of the mean occurred. In these six distributions, the x-axis is values of the mean; the y-axis is frequency.

First, look at the normal distribution, with its three sampling distributions below. When we sample a normal population, the central limit theorem says (see above) that the sampling distribution should be normal. Notice that we have a normal distribution for n=2, 5, and 30. The distributions are not perfectly normal, especially for n=2 and n=5. Remember that the computer actually did this - this is a real exercise - not just a theoretical presentation. Therefore, the distributions are not perfect - but they are statistically normal.

Now, let's consider the uniform distribution. When we sample from the uniform distribution, we are certainly sampling a nonnormal population. Remember what the central limit theorem says (see above), i.e. when sampling a nonnormal population that your sampling distribution should approach normality, getting closer for larger sample sizes. Look at the sampling distributions below the uniform population (remember - these show how often different values of the **mean** occurred). Notice that the sampling distributions in the figures become more normal as sample size goes from 2 to 5, and from 5 to 30.

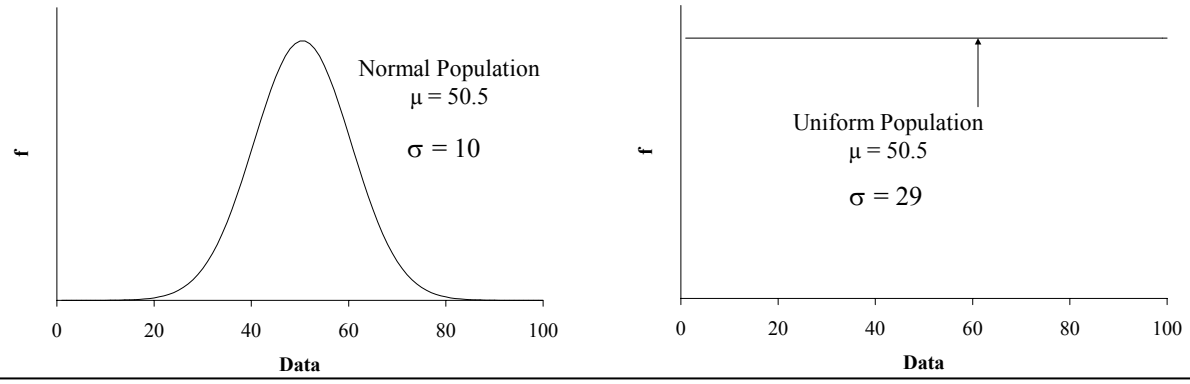
The third thing the central limit theorem tells us is that as sample size increases, that standard error decreases. Standard error is the standard deviation of the sampling distribution. You can see this aspect graphically by looking at the sampling distributions going down in the two columns. Notice that within a column, as you move down the column (i.e. increasing sample size, going from n=2 to n=5 to n=30) that the distributions become "skinnier" (a nonmathematical way to say that the standard deviation is decreasing). The standard error of each sampling distribution is on the graph (standard error is abbreviated as SE on the graphs).

An important property you should notice about the graphs is that the sampling distribution of the mean for both populations is essentially normal at the bottom, where n=30. This means that a sample size of 30 yields a fairly accurate estimate of the mean, regardless of what kind of distribution you are sampling. The sampling distribution of means taken from the uniform distribution is "just as normal" as the sampling distribution of means from a normal population. With n=30 we can be reasonably certain that our mean has a normal sampling distribution.

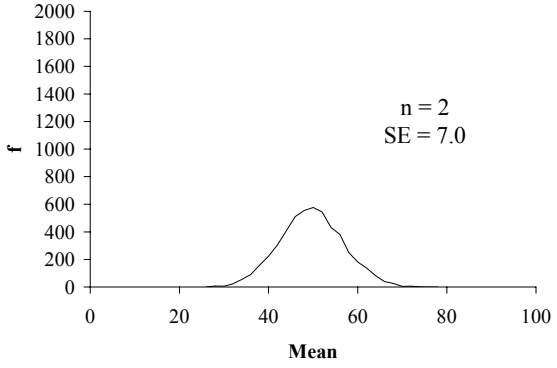
This is the reason why statisticians consider the cutoff between "large" and "small" samples to be right around n=30. Statisticians know that a random sample of size n=30 yields a pretty accurate estimate of the mean, regardless of what the population being sampled looks like. Most people are surprised to learn that 30 constitutes a "large" sample. The popular belief that large samples must be in the hundreds, or thousands, or even millions is simply not supported by the central limit theorem.

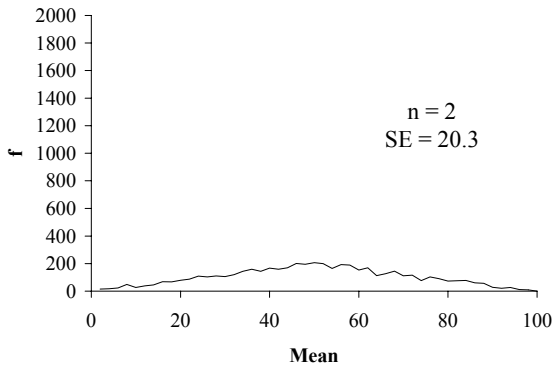
Finally, you might have noticed that the standard error of the sampling distribution from the uniform distribution at n=30 is larger than the standard error of the sampling distribution from the normal distribution at n=30. That is, look at the two distributions at the bottom of the illustration, notice that the one on the right is "fatter", i.e. more variable. This is is because the populations being sampled (see the top of the illustration) have very different standard deviations- remember the $\sigma = 10$ for the normal and $\sigma = 29$ for the uniform. This is another important lesson to learn: **the amount of error involved in estimating a mean increases with larger population variability**.

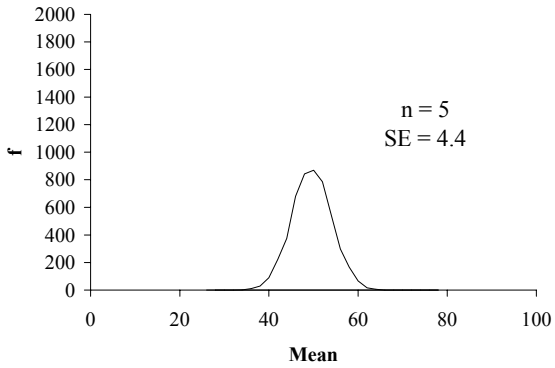
Illustration for the Central Limit Theorem

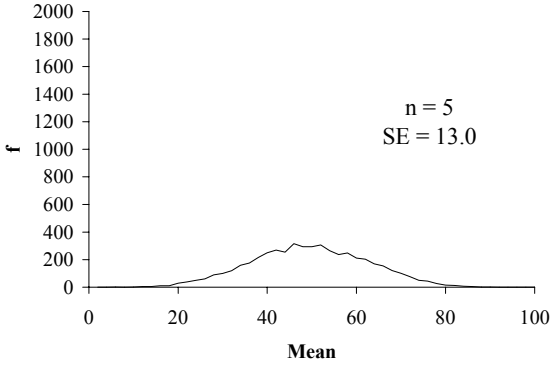


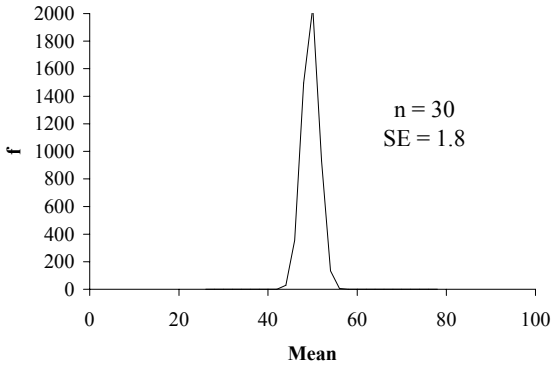
The distributions below are sampling distributions of the mean.

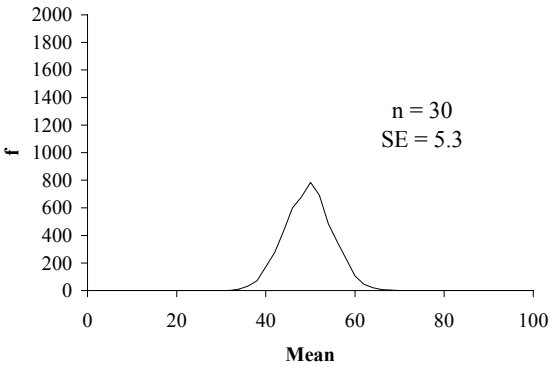












PRACTICE EXAMS

The following pages contain four practice tests for the open book portion of the exam. Tests 1A and 1B are practice for the first midterm; and test 2A and 2B are practice for the second midterm. The final exam is comprehensive, so both practice exams can be used to help prepare for the final exam. A few sample questions of the type you will see on the closed book portion of the exams are also provided. The answers for all questions are also given.

The open book practice exams are exact simulations of the real exams. The number of questions, distribution of points, and general types of questions are exactly what you will see on the actual exams.

It is highly recommended that you use the practice tests in an actual simulation of the exam. That is, try doing the Open Book section in about 35 minutes. This gives the best preparation in that it puts you under some time pressure to answer the questions quickly and accurately. Try to avoid reading a question and simply turning back to the answer sheet to find the solution; this will help you very little.

A difficult part of the exams (and of actually doing data analysis) is determining which test is appropriate to use. Some problems on the open book portion of the exams give you data and a biological question, and you must determine the test to be used. The practice exams are good practice in choosing the proper analysis before having to do it on an actual exam. Doing problems at the end of the chapters in Zar helps familiarize you with the mechanics of doing analyses, but picking which test to use is easy. You simply use the procedures that were discussed in the particular chapter. On the practice tests (just like on real tests or in real life), problems don't occur at the end of chapters. You must have a systematic procedure to allow you to determine what analysis is appropriate by using the biological question and data. Be certain to use the practice exams to help you develop and refine your procedure.

Additional practice in choosing the correct test is available after the practice tests in the section on the Werner Blood Chemistry Data problems. You should do these problems before the final exam (not before the second exam).

CALIFORNIA STATE POLYTECHNIC UNIVERSITY, POMONA

Biology 211 EXAM 1 Closed Book section. 1 point each. Use SCANTRON form 882.

1. What data scale is appropriate in a Contingency Table Analysis?
   A) Normal.
   B) Ordinal.
   C) Nominal.
   D) Ratio-Interval.
   E) Periodic.

2. A datum is
   A) a summary of central tendency and variability in a sample.
   B) a numerical feature of a sample.
   C) a numerical feature of a population.
   D) another term for statistic.
   E) a numerical fact.

3. What is (are) the assumption(s) of the One-sample t-test with respect to the distribution of the data?
   A) The differences are a random sample from a normally distributed population of differences.
   B) All samples are random samples from normally distributed populations.
   C) None.
   D) Samples are random samples from a bivariate normal distribution.
   E) All samples are random samples from normally distributed populations with equal variances.

.....***you should expect about 20 questions of this type***.....

20. What is tested in a Goodness-of-Fit Test?
   A) To see if observed frequencies and *a posteriori* expected frequencies are the same.
   B) Difference between a set of observed frequencies and a set of expected frequencies generated from observed ratios.
   C) To see if the data accurately estimate the selected statistical model.
   D) Difference between a set of observed frequencies and a set of expected frequencies generated by an *a priori* ratio or distribution.
   E) To see if the frequencies of one variable are different from the frequencies of a second variable.

CALIFORNIA STATE POLYTECHNIC UNIVERSITY, POMONA

Biology 211 EXAM 1A Name_____ID#_____

Open Book Questions. Show answer and method in the space provided. 16 points each.

1. (16 points) 1000 chi-squared values (with 1 degree of freedom) have been calculated on 1000 independent samples where the null hypothesis is true. How many of these chi-squared values would you expect to be less than 0.102?

2. (16 points) 100 dogs of a particular breed were found to average 16.4 pounds (standard deviation = 1.44) in body weight. The American Kennel Club (AKC) states that this breed should average 14 pounds. Test to determine if the sample differs significantly from the AKC weight. Assume relevant distributions are normal.

3. (16 points) An ornithologist noted in a recent publication that the extinct Knucklehead Finch (*Condyla cephala*) used to occur in three color morphs: red, purple, and mauve. He also stated that mauve was twice as abundant as either red or purple. I have discovered a previously unknown collection of Knucklehead Finches in the basement of the old science building. This collection contains 16 mauve, 6 purple, and 2 red birds. Test the data from this collection relative to the published statements.

-Exam 1A Page 2-

4. (16 points) Test for a decrease in the percentage of plants infested by insects on the plots after treatment. Assume relevant distributions are nonnormal.

	Percent Plants Infested	
Plot	Before Treatment	After Treatment
1	27.0	21.0
2	22.0	19.0
3	19.0	21.0
4	20.0	20.5
5	23.0	22.0
6	29.0	24.0
7	33.0	26.0
8	25.0	21.0

5. (16 points) Below are the scores on a biometrics exam from a previous year. Scores are expressed as percentages. Test for a difference between males and females in performance on the exam. Percentages are not normally distributed.

Male	Female
87	85
57	62
45	62
73	75
33	65
16	70
35	68
60	37
	64

CALIFORNIA STATE POLYTECHNIC UNIVERSITY, POMONA

Biology 211 EXAM 1B Name_____ID#_____

Open Book Questions. Show answer and method in the space provided. 16 points each.

1. If a contingency table with 5 rows and 6 columns is tested using a sample of data from a population where the null hypothesis of independence is known to be true, what is the probability that the test statistic calculated will be greater than or equal to 10.851?

2. Test to determine if the letters A, B, and C occur with equal frequency in the population from which the following list was sampled:

B C B A C C B C B A B C B C B A B C A C C B C B

3. Test for a difference in the wing length of butterflies between California and Texas. Assume that the relevant distributions are not normal. Data are in millimeters.

California	Texas
41	48
44	43
42	50
46	45
47	49
40	51

BIO 211 Biometrics Exam 1B OPEN BOOK Section Page 2

4. Test to determine if 25 lizards averaging 20 cm in length (s^2=4) came from a population averaging 20.9 cm. Assume relevant distributions are normal.

5. A study was conducted to investigate whether oat bran cereal helps to lower serum cholesterol levels. Fourteen subjects were given a diet containing corn flakes, and after two weeks their cholesterol levels were measured (data below). The same subjects were then given a diet containing oat bran, and after two weeks their cholesterol levels were again measured (data below). Test for a difference in the variability of cholesterol level with the corn flakes versus the oat bran diet. Assume all relevant distributions to be normal.

	Cholesterol Level (mmol/l)	
Subject	Corn Flakes	Oat Bran
1	4.61	3.84
2	6.42	5.57
3	5.40	5.85
4	4.54	4.80
5	3.98	3.68
6	3.82	2.96
7	5.01	4.41
8	4.34	3.72
9	3.80	3.49
10	4.56	3.84
11	5.35	5.26
12	3.89	3.73
13	2.25	1.84
14	4.24	4.14
Mean	4.444	4.081
Variance	0.939	1.117

BIO 211 Practice Exam Answers

Exam 1 - Closed Book section. (Multiple Choice Questions)

1. C

2. E

3. B

.....

20. D

BIO 211 Practice Exam Answers

Exam 1A - Open Book Section

1. $P(\chi^2 \geq 0.102) = 0.75$. Therefore, the probability that chi-squared will be less than 0.102 is 1-0.75, or 0.25. Since there were 1000 values calculated, we expect 250 (1000*0.25) to be less than 0.102.

2. Ho: $\mu = 14$
The standard error is the standard deviation (1.44) divided by the square root of the sample size (n = 100), or 1.44 / 10 = 0.144.

$$t = \frac{16.4 - 14}{0.144} = 16.667$$

DF = n - 1 = 100 - 1 = 99 $p < 0.05$ Reject Ho: Notice that although 99 degrees of freedom does not appear in the t-table, that since we could safely reject Ho: with 98 degrees of freedom (which is in the table), we can surely reject with 99 degrees of freedom.

3. Ho: No difference between the observed frequencies and those expected from a 2:1:1 (Mauve:purple:red) ratio. Let f represent observed frequency, and F represent expected frequency.

	f	F	$(f-F)^2/F$	
mauve	16	12	1.33	
purple	6	6	0	DF = k - 1 = 3 - 1 = 2
red	2	6	2.67	$p > 0.05$ Accept Ho:
			4.00	$\chi^2 = 4.00$

4. Wilcoxon Paired-sample test, one-tailed. Ho: % plants infested before treatment is less than or equal to % plants infested after treatment. See Zar for procedure (4th edition: page 168; 5th edition: page 186). Referring to Zar, page 166, we see that for our null hypothesis, we should use T_- as the test statistic. The sum of the negative ranks is: $T_- = 3 + 1 = 4$. A Wilcoxon T of 4 with n=8 has an alpha probability of $p < 0.05$, therefore Ho: is rejected.

5. Mann-Whitney U. Ho: No difference between males and females in test scores. Ranking from high to low, the ranks for the males are (starting with 87): 1, 12, 13, 4, 16, 17, 15, 11. Sum of these ranks is 89. Ranks for the females (starting with 85): 2, 9.5, 9.5, 3, 7, 5, 6, 14, 8. Sum of these ranks is 64. Calculated U is 19, and U'is 53, therefore 53 is used as the test statistic. The sample sizes are 8 (males) and 9 (females); $p > 0.05$; Accept Ho:.

Notice that in the original data, there is a large difference in mean (females = 65.3; males = 50.8). But even if we do the more powerful Two-sample *t*-test, we find no difference. This is due to the large variability within the groups (male scores go from 33 to 87; females from 37 to 85), therefore the 15 point mean difference is not significant.

BIO 211 Practice Exam Answers

Exam 1B - Open Book Section

1. This is a "table look-up" problem. The probability we are asked to determine (i.e. $P(\chi^2 \geq 10.851)$) is the alpha probability. We need the degrees of freedom in order to look-up the probability. In a contingency table, DF=(r-1)(c-1), so our DF=(5-1)(6-1)=(4)(5)=20. Looking in Zar's Table B.1 (4th edition: page App12; 5th edtion: page 672) we find the row where DF=20, and note that the alpha probability associated with 10.851 is 0.95, so that's our answer: p = 0.95.

2. Chi-squared Goodness-of-fit. If A, B, and C occur with equal frequency, we expect them to have a 1:1:1 ratio. In the row of letters, we observe 4 A's, 10 B's, and 10 C's. Ho: no difference between the observed frequencies of A, B, and C, and those expected from a 1:1:1 ratio.

	f_i	$\hat{f}_i$	$\frac{(f_i-\hat{f}_i)^2}{\hat{f}_i}$
A	4	8	2.0
B	10	8	0.5
C	10	8	0.5
	24	24	3.0

Since there are 24 total letters, we expect 1/3 of them (i.e. 8) to be A, 1/3 to be B, and 1/3 to be C.

DF = k-1 = 3-1 = 2 $\chi^2 = 3.0$

p>0.05 Accept Ho:

3. Mann-Whitney U-test. There is no biological relationship between a particular butterfly from California and one from Texas, so there is no way to pair these data. Each butterfly in California is independent of each butterfly in Texas. Since distributions are nonnormal, a nonparametric test should be used.

California	Rank	Texas	Rank
41	2	48	9
44	5	43	4
42	3	50	11
46	7	45	6
47	8	49	10
40	1	51	12
	---		---
$R_1 = 26$		$R_2 = 52$	

Ho: No difference in wing length between CA and TX

$$U = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 = 6\times 6 + \frac{6(6+1)}{2} - 26 = 31$$

$$U' = n_1 n_2 - U = 6\times 6 - 31 = 5$$

Since this is a two-tailed test, we use the larger of U or U'; so out test statistic is U=31.

In the table (B.11, 4th edition: App92; 5th edition: page 750) we see that the critcal value at $\alpha(2)=0.05$ is 31, so our p=0.05.

We reject Ho:

4. One-sample *t*-test. Since the variance of the 25 lengths is 4 ($s^2 = 4$), the standard deviation is 2 (s = 2).

$$s_{\overline{X}} = \frac{s}{\sqrt{n}} = \frac{2}{\sqrt{25}} = 0.4$$

$$t = \frac{\overline{X}-\mu}{s_{\overline{X}}} = \frac{20-20.9}{0.4} = -2.25$$

Ho: $\mu = 20.9$

DF = n - 1 = 25 - 1 = 24

p < 0.05 Reject Ho:

Practice Exam 1B - Answers to Open Book section (continued)

5. Since we are asked to test for a difference in variability, the Variance Ratio Test is appropriate. There is no *a priori* expectation of which variance should be larger, so this a two-tailed test, and we divide the larger variance by the smaller.

$H_O: \ \sigma_1^2 = \sigma_2^2$

$F = \frac{s_2^2}{s_1^2} = \frac{1.117}{0.939} = 1.19$

$DF_{numerator} = n_2 - 1 = 14 - 1 = 13$

$DF_{denominator} = n_1 - 1 = 14 - 1 = 13$

$p > 0.05$ Accept Ho:

CALIFORNIA STATE POLYTECHNIC UNIVERSITY, POMONA

Biology 211 EXAM 2 Closed Book section. 1 point each. Use SCANTRON form 882.

1. What data scale is appropriate in a Kruskal-Wallis Test?
   A) Normal.
   B) Ordinal.
   C) Nominal.
   D) Ratio-Interval.
   E) Periodic.

2. A level is
   A) a categorical variable used in an ANOVA.
   B) a variable being measured by each datum in an ANOVA.
   C) a category of a grouping variable in an ANOVA.
   D) a mean of a cell in an ANOVA.
   E) a mean of all the data points in an ANOVA.

3. What is (are) the assumption(s) of One-factor ANOVA with respect to the distribution of the data?
   A) The differences are a random sample from a normally distributed population of differences.
   B) All samples are random samples from normally distributed populations.
   C) None.
   D) Samples are random samples from a bivariate normal distribution.
   E) All samples are random samples from normally distributed populations with equal variances.

.....***you should expect about 20 questions of this type***.....

20. What is tested in a Regression analysis?
   A) To see if observed frequencies and *a posteriori* expected frequencies are the same.
   B) Difference between two sample means.
   C) To see if their is a linear relationship between two variables.
   D) To see if data points tend to revert back towards an *a priori* mean.
   E) Equality of 3 or more variances (homoscedasticity).

CALIFORNIA STATE POLYTECHNIC UNIVERSITY, POMONA

Biology 211 EXAM 2A Name_____ID#_____

Open Book Questions. Show answer and method in the space provided. 16 points each.

1. (16 points) If an analysis-of-variance is performed on five groups of ten data points each (all samples are from the same population), what is the probability of calculating an F less than 3.09?

2. (16 points) Four samples (each of size n=5) of femur lengths from four different subspecies of the ermine (Mustela erminea) are taken. The variances of the four samples are: 2.0, 2.5, 3.0, and 3.5. The variance of the mean of each sample about the mean of all 20 data points is 5.5. Test to determine if the average femur length is the same for all the subspecies. Present your results in the appropriate table. Assume all samples are normally distributed and have equal variances.

3. (16 points) In the leopard frog (Rana pipiens), 50% of the variation in heart rate can be explained by regression on ambient temperature. The variance of heart rate is 6.0 (n=12). Test this regression for significance using the analysis of variance technique. Present your results in the proper table.

-Exam 2A Page 2-

4. (16 points) The literature suggests that larger individuals of a certain species of bird have a higher percentage of insects in the diet than smaller individuals. Test this suggestion nonparametrically with the following data. Percentages are not normally distributed.

Body Weight (g)	% Insects in Diet
59	38
60	35
57	32
58	28
61	39
62	39

5. (16 points) Four subspecies (designated A, B, C, and D) of a desert shrub are tested for salt tolerance by growing plants from seeds in 0.3% salt solution. Ten different greenhouses (with different light and temperature control) were used. One plant from each of the four subspecies was grown in each greenhouse. After a set length of time, growth was determined by oven-dry weight of the plants. The data (in grams) are given below. Also given are the mean weights for each greenhouse and for each subspecies (n=10). Test for: 1) a difference among the subspecies; and 2) a difference among the greenhouses. Complete the table below as part of your tests, and be sure to state your null hypotheses. Assume relevant distributions are normal and homoscedastic.

Greenhouse	A	B	C	D	Mean (n = 4)
1	1.28	1.41	1.40	1.43	1.38
2	1.43	1.62	1.56	1.69	1.58
3	1.29	1.49	1.43	1.42	1.41
4	1.43	1.68	1.59	1.62	1.58
5	1.26	1.40	1.46	1.47	1.40
6	1.39	1.56	1.58	1.56	1.52
7	1.31	1.55	1.47	1.47	1.45
8	1.37	1.60	1.51	1.51	1.50
9	1.38	1.54	1.44	1.58	1.49
10	1.27	1.44	1.42	1.44	1.39
Mean (n=10)	1.34	1.53	1.49	1.52	

Source	SS	DF	MS
Total	.462		
Subspecies			
Greenhouses			
Error	.033		

CALIFORNIA STATE POLYTECHNIC UNIVERSITY, POMONA

Biology 211 EXAM 2B Name_____ID#_____

Open Book Questions. Show answer and method in the space provided. 16 points each.

1. What is the two-tailed probability that no linear relationship exists between two variables where 9% of the variation in one variable is explained by regression on the other variable using 60 data points?

2. An immunologic evaluation is made on eleven patients with Kaposi's sarcoma, and levels of serum immunoglobulins G (IgG) and A (IgA) are determined. Test for an association between IgG and IgA levels. Data are mg/dl. Assume relevant distributions are not normal.

Patient	IgG	IgA
1	1262	242
2	1137	154
3	1118	113
4	1882	152
5	2777	822
6	2483	872
7	1335	380
8	2757	215
9	1900	365
10	1466	264
11	1518	503

3. Four mice each receive three treatments (each treatment on a different day), and are then timed for how long it takes the mouse to find its way through a maze. The data are analyzed with a randomized block ANOVA. Complete the ANOVA table below, and (assuming no interaction) state and test all possible null hypotheses.

Source	SS	DF	MS
Total			
Treatments (groups)	44	2	
Mice (blocks)	33		
Error (remainder)	1		

-Exam 2B Page 2-

4. A Two-factor Analysis of Variance (Model I) with replications is performed. The data consist of three levels in one factor and four levels in the second factor; with three replications in each cell. In testing the hypothesis of no interaction, what is the probability of obtaining a value of the test statistic greater than or equal to 4.20 just due to random sampling error?

5. The average number of eggs per nest in the nests of 40 California Quail (Callipepla californica) was 8.3 (standard deviation = 2.0). The sample of 40 nests consists of 10 nests from each of four counties. The variances of the eggs per nest for each county are given below. Test to determine whether the mean number of eggs per nest is the same for the four counties. Assume homoscedasticity and that relevant distributions are normal. Present your results in the appropriate table.

County	Variance
Los Angeles	1.0
San Bernardino	1.5
Riverside	1.0
San Diego	1.5

BIO 211 Practice Exam Answers

Exam 2 - Closed Book Section (Multiple Choice Questions)

1. B.

2. C.

3. E.

.....

20. C.

BIO 211 Practice Exam Answers

Exam 2A - Open Book Section

1. Groups DF = k -1 = 5 - 1 = 4; Error DF = N - k = 50 - 5 = 45. The one-tailed probability of F being greater than or equal to 3.09 with 4 and 45 DF is 0.025. Therefore, the probability of F being less than 3.09 is 0.975.

2. ANOVA, Ho: $\mu_1 = \mu_2 = \mu_3 = \mu_4$. Since you know there are 4 groups of n=5, there must be 20 total data points. All degrees of freedom can then be calculated. You must recognize from the description that 5.5 is GROUPS MS. The really tricky part is to realize that since you know the variance of each group, and the sample size of each group, that you can calculate the sum of squares (SS) of each group as follows: since variance = SS / n-1, therefore SS = variance * (n - 1). SS quantities for each group are: 2.0 * 4 = 8, 2.5 * 4 = 10, 3.0 * 4 = 12, 3.5 * 4 = 14. ERROR SS is the pooled SS for each of the groups, or 8 + 10 + 12 + 14 = 44. Now that you have this quantity, you can calculate the rest of the table.

SOURCE	SS	DF	MS	
Total	60.5	19	3.2	
Groups	16.5	3	5.5	F = 5.5 / 2.75 = 2.00
Error	44	16	2.75	p > 0.05 Accept Ho:

If you got this one right, you are doing WELL! Study the sources of variation!

3. Ho: ß = 0. Regression DF = 1, Residual = n - 2 = 12 - 2 = 10. Since heart rate is the dependent variable and Total MS is the variance of the dependent variable, Total MS must be 6.0, and Total SS must be 6.0 * 11 = 66. Coefficient of determination is 0.50, and coefficient of determination = Regression SS / Total SS. Therefore, Regression SS = (coefficient of determination) * (Total SS) = 0.5 * 66 = 33. Now, the entire table can be completed:

Source	SS	DF	MS	
Total	66	11	6	F = Regression MS / Residual MS
Regression	33	1	33	= 33 / 3.3 = 10.0
Residual	33	10	3.3	p < 0.05 Reject Ho:

If you got this one, you are doing great! Study the sources of variation!

4. Spearman rank correlation. Ho: $\rho_s \leq 0$. This is one-tailed.

Rank of body weight	Rank of %Insects	di	di^2
3	4	-1	1
4	3	1	1
1	2	-1	1
2	1	1	1
5	5.5	- .5	.25
6	5.5	.5	.25
		Σdi^2 = 4.5	

$r_s = 1 - 27 / (216 - 6) = 0.871$. $p < 0.05$; Reject Ho:

Practice Exam 2A - Answers to Open Book section (continued)

5. Two-factor ANOVA (no replications, randomized block design). Although there are several ways to approach this problem (including doing the whole analysis from the beginning), the easiest method is to recognize that if you calculate either Subspecies SS or Greenhouse SS, you can do the analyses. You should be able to fill in the degrees of freedom immediately. Subspecies SS is the easiest to calculate. Subspecies SS is a Groups SS and is equal to the sum of each subspecies sample size (i.e. 10) times the squared deviation between the subspecies mean and the grand mean. Or, in terms of a formula:

$$Groups\ SS = \sum_{j=1}^{4} n_j (\overline{X}_j - \overline{X})^2$$

In order to do this calculation, you need to know the grand mean (the mean of all 40 data points). You can just do this calculation, but it is quicker to realize that since each subspecies has the same sample size (10), that the grand mean is the mean of the subspecies means, i.e. (1.34 + 1.53 + 1.49 + 1.52) / 4 = 1.47. You now use the formula for Subspecies SS:

$10 * (1.34 - 1.47)^2 +$

$10 * (1.53 - 1.47)^2 +$

$10 * (1.49 - 1.47)^2 +$

$10 * (1.52 - 1.47)^2 = .169 + .036 + .004 + .025 = .234 = \text{Subspecies SS}$

You can now complete the entire table and perform the two tests:

Source	SS	DF	MS
Total	.462	39	
Subspecies	.234	3	.078
Greenhouses	.195	9	.022
Error	.033	27	.001

Testing for a difference among subspecies: Ho: $\mu_1 = \mu_2 = \mu_3 = \mu_4$
F = .078 / .001 = 78 p < .05 Reject Ho:

Testing for a difference among greenhouses: Ho: $\mu_1 = \mu_2$ ..... $= \mu_{10}$
F = .022 / .001 = 22 p < .05 Reject Ho:

You cannot test for interaction because there are no replications. If an interaction between greenhouses and subspecies is suspected, then the test for equality of greenhouse means is not appropriate.

Again, this is a hard problem. Congratulations if you solved it!

STUDY THE SOURCES OF VARIATION!!!

BIO 211 Practice Exam Answers

Exam 2B - Open Book Section

1. There are two things you need to know here: (1) the hypothesis of no linear relationship can be tested by either regression (Ho: β=0) or correlation (Ho: ρ=0); and (2) that the 9% is the coefficient of determination (r^2). Since we are told r^2, we can calculate the correlation coefficient (r) by taking the square root of 0.09, which is 0.3. Therefore, r = 0.3. DF = n-2 = 60-2 = 58. We look in Zar Table B.17 (4th edition: page App109; 5th edition: page 766) and find that the α(2) probability of an r = 0.3 with 58 degrees of freedom is 0.02. Our answer is p = 0.02.

2. We are asked to test for an association between IgG and IgA levels, so we want to do a test in the regression/correlation family of tests. Since we have nonnormal distributions, we should do a nonparametric test. Therefore, we must do a Spearman Rank Correlation test. Ho: $\rho_s = 0$

Patient	IgG	Rank	IgA	Rank	d_i	d_i^2
1	1262	3	242	5	-2	4
2	1137	2	154	3	-1	1
3	1118	1	113	1	0	0
4	1882	7	152	2	5	25
5	2777	11	822	10	1	1
6	2483	9	872	11	-2	4
7	1335	4	380	8	-4	16
8	2757	10	215	4	6	36
9	1900	8	365	7	1	1
10	1466	5	264	6	-1	1
11	1518	6	503	9	-3	9
						98

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n} = 1 - \frac{6 \times 98}{11^3 - 11} = 1 - \frac{588}{1331 - 11} = 1 - \frac{588}{1320} = 0.555$$

n = 11. Checking Table B.20 (4th edition: page App116; 5th edition: page 773) we find p > 0.05. Accept Ho:

3. Total SS is found by adding together all of the other SS. The formulas for the DF values are in Zar, appear earlier in this TestPac, and should be in your lecture notes. The MS values are calculated by dividing the SS values by the DF values. Since we are assuming no interaction we can test equality of the block means in addition to equality of the treatment means.

Source	SS	DF	MS	Ho:	F	p	Conclusion
Total	78	11					
Treatments (Groups)	44	2	22	$\mu_1=\mu_2=\mu_3$	131.7	<0.05	Reject Ho:
Mice (Blocks)	33	3	11	$\mu_1=\mu_2=\mu_3=\mu_4$	65.9	<0.05	Reject Ho:
Error (Remainder)	1	6	0.167				

Remember that the F values are calculated as the Factor MS / Error MS.
Treatments F = 22 / 0.167 = 131.7 Blocks F = 11 / 0.167 = 65.9

Practice Exam 2B - Answers to Open Book section (continued)

4. To answer the question, you only need to know the degrees of freedom for interaction and error. To get this, try to visualize how the data are structured. Let each * represent a datum.

		Factor 1		
		Level 1	Level 2	Level 3
Factor 2	Level 1	* * *	* * *	* * *
	Level 2	* * *	* * *	* * *
	Level 3	* * *	* * *	* * *
	Level 4	* * *	* * *	* * *

There are 36 data points, so Total DF = N-1 = 36-1 = 35.
Factor 1 has 3 levels, so Factor 1 DF = levels - 1 = 3-1 = 2
Factor 2 has 4 levels, so Factor 2 DF = levels - 1 = 4-1 = 3
Interaction DF is the product of the DF for the two factors = 2 × 3 = 6
We can calculate Error DF by subtraction = 35 - 2 - 3 - 6 = 24

Another method to calculate Error DF is to see that each cell has 3 data points. Therefore each cell has 2 DF. If we pool (add up) the DF for each of the 12 cells, we get 2+2+2+2+2+2+2+2+2+2+2+2 = 12 x 2 = 24.

Since the test for interaction is Interaction MS / Error MS, we need the $\alpha(1)$ probability of F=4.2 with 6 DF in the numerator and 24 DF in the denominator. We look at Table B.4 (4^{th} edition: page App26; 5^{th} edition: page and see that the probability is 0.005.

5. This is a One-factor ANOVA. The easiest things to get are always the DF values, so let's do that first. Total DF is N-1. There are 40 nests, Total DF = 40 - 1 = 39. The factor (County) has 4 levels, and the DF for a factor (i.e. Groups DF) is the number of levels -1, so Groups DF = 3. Error DF can most easily be done by subtraction: Error DF = 39 - 3 = 36.

   The key for the rest of the problem is understanding our sources of variation! We are told the standard deviation of all the data points is 2.0, so the variance of all the data points must be $2^2 = 4$. The variance of all the data points is Total MS, so Total MS = 4. We are given the variance for each group, and we know that each group has an n=10, so we can calculate the SS for each group: $SS = s^2(n-1)$. We then pool them (add them) to get Error SS

County	Variance		n-1		SS
Los Angeles	1.0	×	9	=	9.0
San Bernardino	1.5	×	9	=	13.5
Riverside	1.0	×	9	=	9.0
San Diego	1.5	×	9	=	13.5

					Error SS = 45

Continued on the next page ↓

Practice Exam 2B - Answers to Open Book section (continued)

5. (continued)

We can now construct the ANOVA table and test the null hypothesis.....

Ho: $\mu_1=\mu_2=\mu_3=\mu_4$

Source	SS	DF	MS
Total	156	39	4
Groups	111	3	37
Error	45	36	1.25

F = 37 / 1.25 = 29.6 p < 0.05 Reject Ho:

This problem is an excellent example of the "puzzles" you'll have to solve on the actual exams. The key to the puzzle is understanding your ANOVA sources of variation.

Study the sources of variation!

PRACTICE PROBLEMS - WERNER BLOOD CHEMISTRY DATA

One of the most difficult aspects of biometrics is choosing the correct test to perform given some data and a biological question. The purpose of these pages is to give you some practice in choosing the correct test and stating the correct null hypothesis.

Below you will find a small subset of the Werner et al.(1970) blood chemistry data. Following the data are a series of biological questions. For each biological question, determine 1) the correct statistical test; and 2) state the null hypothesis. The answers are given on a separate page - but DO NOT CONSULT THE ANSWERS FIRST! The exercise loses its value if you look at the answers before you decide which test should be done and how the null hypothesis should be stated.

The problem set has questions covering most of the statistical tests from the course. Remember that the second hourly exam is not comprehensive, but the final exam is.

This exercise should also demonstrate to you how many different and varied biological questions can be asked of a single data set. Biometrics provides a necessary tool to help search for answers to these questions.

THE WERNER BLOOD CHEMISTRY DATA

In order to choose the correct statistical test, you must be familiar with the data. Here the data variables are described. Read this section carefully, and refer back to it when you are choosing tests.

The Werner blood chemistry data set consists of physical and blood chemistry measurements on women who have been matched by age and by their status relative to taking contraceptive pills. Each line of data represents a different woman. The 16 lines of data thus represent 16 women. The 16 women are matched by age into 8 sets. In each set of two matched women, one of them uses oral contraceptives while the other does not.

Werner, M., R. Tolls, J. Hultin, and J. Mellecker. 1970. Sex and age dependence of serum calcium, inorganic phosphorus, total protein, and albumin in a large ambulatory population. In *Fifth International Congress on Automation, Advances in Automated Analysis*, Vol. 2, 59-65.

The variables in the data are:

ID Number - this is not really a measurement, but just a number to identify individuals. Women matched by age are always on successive lines.

Age - In years.

Height - In inches.

Weight - In pounds.

Pill - A 1 indicates that the woman does not use the contraceptive pill; a 2 means the pill is taken.

Chol - Amount of cholesterol in the blood. Units are mg/dl.

Album - Amount of albumin in the blood. Units are mg/dl.

Calc - Amount of calcium in the blood. Units are mg/dl.

UrAc - Amount of uric acid in the blood. Units are mg/dl.

ID Number	Age	Height	Weight	Pill	Chol	Album	Calc	UrAc
2381	22	67	144	1	200	4.3	9.8	5.4
1946	22	64	160	2	220	3.5	8.8	7.2
1610	25	62	128	1	243	4.1	10.4	3.3
1797	25	68	150	2	265	3.8	9.6	3.0
561	19	64	125	1	158	4.1	9.9	4.7
2519	19	67	130	2	255	4.5	10.5	8.3
225	20	64	118	1	210	3.9	9.5	4.0
2420	20	65	119	2	192	3.8	9.3	5.0
1649	21	60	107	1	246	4.2	10.1	5.2
3108	21	65	135	2	245	3.4	10.6	4.8
2698	26	66	135	1	240	4.8	10.3	5.1
3006	26	64	118	2	238	4.0	9.9	4.6
883	27	63	125	1	168	4.2	9.7	4.1
1882	27	64	124	2	200	4.0	9.6	5.2
609	30	64	135	1	174	4.0	9.5	3.5
3021	30	66	112	2	250	4.4	10.0	3.5

Biological Questions:

1. Is there a difference in cholesterol between women using the pill and those not? Assume the relevant distributions to be normal.

2. For only those women using the pill, is cholesterol amount associated with uric acid amount? Assume the relevant distributions are not normal.

3. Dividing the women into those over 125 pounds and those under or equal to 125, is the proportion of women on the pill the same for these two groups? Assume the relevant distributions are normal.

4. Using the same groups as in number 3 (>125 and #125), does the heavier group have more uric acid in the blood? Disregard pill status and age. Assume the relevant distributions are normal.

5. Do women on the pill have lower calcium amounts in the blood? Assume the relevant distributions are not normal.

6. Are women on the pill more variable in albumin content than those not on the pill? Assume the relevant distributions are normal.

7. According to the National Inquirer, women on the pill average five and one-half feet tall. Is this true for our data? Assume the relevant distributions are normal.

8. Disregarding pill status, divide the women into three sets based on age: 19-21, 22-25, and 26-30 years. Are these groups the same in body weight? Assume the relevant distributions are normal.

9. For women using the pill, it has been speculated in the medical literature that heavier body weight is associated with higher blood cholersterol levels. What is the conclusion for our data? Assume the relevant distributions are not normal.

10. Based on his vast knowledge of the base 10 number system, Sir Isaac Newton predicted that a random sample from the Werner data would have ten times as many ID Numbers with 4 digits as ID Numbers with only 3 digits. Is he correct for our data? Assume the relevant distributions are not normal.

11. Are women on the pill heavier than those not on the pill? Assume the relevant distributions are normal.

12. For women not using the pill, is there an association between height and age? Assume the relevant distributions are normal.

13. Disregarding age, is there a difference between women using the pill and those not using the pill in the amount of uric acid in the blood? Assume the relevant distributions are not normal.

14. Same question as in #13 above, except do not disregard age. Assume the relevant distributions are normal.

15. Divide the data into three groups on the basis of age just like in #8 above (i.e. 19-21, 22-25, and 26-30 years). Is there any difference among these three groups in cholesterol level? Assume the relevant distributions are normal.

ANSWERS

1. Paired-sample *t*-test. Ho: $\mu_d = 0$. Women are paired by age.

2. Spearman rank correlation. Ho: $\rho_s = 0$

3. Contingency table analysis. Ho: Relative frequency of lighter group is independent of pill status

4. Two-sample *t*-test. If group 1 is >125, then Ho: $\mu_1 \leq \mu_2$

5. Wilcoxon paired-sample test. Ho: On pill $\geq$ not on pill. Women are paired by age.

6. Variance ratio test. If group 1 is not on pill, then $Ho: \sigma_1^2 \geq \sigma_2^2$ It's one tailed because you are asked if women on the pill are more variable.

7. One-sample *t*-test. Ho: $\mu = 66$

8. One-factor ANOVA. Ho: $\mu_1 = \mu_2 = \mu_3$

9. Spearman Rank Correlation. Ho: $\rho_s \leq 0$ It's one tailed because you're testing to see if heavier women have higher cholesterol levels.

10. Goodness-of-fit test. Ho: No difference between the observed frequencies and those expected if the 4-digit:3-digit ratio is 10:1

11. Paired-sample *t*-test. If group 1 was not on pill and group 2 was on pill, so differences are calculated (not on pill - on pill) then Ho: $\mu_d \geq 0$. Women are paired by age.

12. Correlation. Ho: $\rho = 0$.

13. Mann-Whitney U test. Ho: No difference between women using the pill and women not using the pill in uric acid level.

14. Paired-sample *t*-test. Ho: $\mu_d = 0$

15. Two-way analysis of variance, with replication. The factor of interest is age; it has three levels (the three age intervals). The second factor is oral contraceptive status, which has two levels (on pill; not on pill. This is a two-way analysis of variance with replications. It would be necessary and appropriate to test for interaction. If you figured this one out, you are due congratulations.
    Ho: $\mu_1 = \mu_2 = \mu_3$ for the three age groups.
    Ho: $\mu_1 = \mu_2$ for the two contraceptive groups
    Ho: No Interaction

Using a scientific calculator to calculate basic descriptive statistics

There are many scientific calculators available with statistical functions. In order to use your calculator properly, what you should do is **carefully read the manual that came with your calculator**. If you no longer have your manual (and who keeps all that stuff?), you can usually find the manual on the web.

With any model calculator, you must be careful to calculate **sample statistics** and not **population parameters**. In Biometrics, we only calculate sample statistics, e.g. sample variance, sample standard deviation, we do not calculate population parameters such as the population variance or population standard deviation. For example, many calculators represent the sample standard deviation by a symbol something like σ_{xn-1}, while the population standard deviation is represented by σ_{xn}. The "n-1" or the "n" is extremely important. Remember that in the sample variance, the denominator is n-1, while in the population variance, the denominator is n (or N). Therefore, most calculators represent sample values with the "n-1". These are the values you should use. If you are not certain, test with some data where you already know the answer (e.g. from this TestPac, or your text book).

Computer Programs for Statistical Analysis

There is a large number of computer programs available to do statistical analysis on your personal computer. I do not recommend for or against any of these programs. I have never used most of them. I do my statistical computing as follows: (1) for small data sets and simple analyses, I use a spreadsheet (Excel); (2) for large data sets and/or complex analyses, I use SAS.

SPREADSHEETS
Microsoft Excel will do many of the statistical analyses covered in this class. Learning to use a spreadsheet is a wise investment, if you ever think you may have to deal with numbers on any level, not just with statistics. If you use Excel, you'll find statistics under the Tools menu (Data Analysis…). When you go to the Tools menu, if you don't see Data Analysis…, then select "Add Ins..", and then check both Analysis ToolPak and Analysis Toolpak - VBA. Then you should find Data Analysis… under the Tools menu. Spreadsheet programs are available at most software stores. You can get the Microsoft Office programs (including Excel) at the Bronco Bookstore for a special student price.

You are invited to download **StatCat** from Dr. Moriarty's BIO 211 web page. StatCat is an Excel file that makes it easy to do the analyses covered in this class.

THE "BIG PROGRAMS"
There are huge statistical programs that will do virtually everything. They are complex to use. Windows and Macintosh versions are usually available; Windows versions are updated more often. Most universities or research labs have access to one or more of these.

SAS (http://www.sas.com)
Statistical Analysis System. The most widely used major package by biologists. If you decide to learn a big package, pick this one. The California State University has a site license that covers both faculty and students. Go to ehelp for information:
http://www.csupomona.edu/~ehelp/software/download_statistical.shtml

SPSS (http://www.spss.com)
Statistical Package for the Social Sciences. The CSU site license only covers faculty, not students. Purchasing a personal license would be very expensive.

R - The R Project for Statistical Computing (http://www.r-project.org/)
R is a FREE (!) software environment for statistical computing and graphics. It runs on Windows and MacOS. See the web site for information and free download.

Dichotomous Key to the Statistical Tests

Note: this key only includes tests that you have to know how to do. Tests you need to know about (but not actually do) are not included - e.g. Bartlett's Test, Multiple Comparisons, Kruskal-Wallis, and ANCOVA

1a. Data scale is nominal ..... Go to 2
1b. Data scale is ratio, interval, or ordinal ..... Go to 3

2a. One categorical variable, *a priori* ratio or distribution ..... **Goodness-of-fit**
2b. Two categorical variables, Independence ..... **Contingency Table**

3a. One sample (mean vs. *a priori* constant) ..... **One-sample *t*-test**
3b. Two or more samples ..... Go to 4

4a. Multiple samples (more than two), or association between two variables ..... Go to 9
4b. Two samples ..... Go to 5

5a. Testing for difference in variability ..... **Variance Ratio Test**
5b. Testing for difference in central tendency ..... Go to 6

6a. Data samples are independent ..... Go to 7
6b. Data samples are paired ..... Go to 8

7a. Data samples are RSNDP and homoscedastic ..... **Two-sample *t*-test**
7b. Data samples are not RSNDP and/or are not homoscedastic ..... **Mann-Whitney U**

8a. Differences are RSNDP ..... **Paired-sample *t*-test**
8b. Differences are not RSNDP ..... **Wilcoxon paired-sample test**

FIRST EXAM ENDS HERE

SECOND EXAM BEGINS HERE

9a. Difference in central tendency (multiple samples) ..... Go to 10
9b. Association (relationship) between two variables ..... Go to 12

10a. One factor (categorical variable), one response variable ..... **One-factor ANOVA**
10b. Two factors, one response variable ..... Go to 11

11a. One data point per cell ..... **Two-factor ANOVA (Randomized Block)**
11b. More than one data point per cell ..... **Two-factor ANOVA (with replications)**

12a. ANOVA test requested and/or biological causation ..... **Regression**
12b. ANOVA test not requested and/or no causation ..... Go to 13

13a. Bivariate normal distribution ..... **Correlation**
13b. Nonnormal distribution ..... **Spearman Rank Correlation**

Flow Chart of Statistical Tests