

**NOTE: SOME MATERIAL ON TWO-SAMPLE TESTS FOR PROPORTIONS IS FROM CHAPTER 10 IN ROSNER (NOT CHAPTER 8)**

## **PREVIOUSLY**

- estimation

how can a sample be used to estimate the unknown parameters of a population

use confidence intervals around point estimates of central tendency (mean) and variability (variance, standard deviation)

- hypothesis testing, single sample

does a population parameter estimated from sample data differ from some claimed value

almost always results in the same answer as confidence intervals - exception possible with proportions due to difference in how standard errors are calculated in hypothesis testing versus confidence intervals

- in either case, make a statement about a single parameter estimated from sample data

## **NEW QUESTION**

- test a claim about two parameters, both estimated with sample data
- Rosner - underlying parameters of two different populations, neither of whose values are known, are compared
- more frequently encountered than one sample hypothesis testing

Example from Rosner - what is the relationship between oral contraceptive (OC) use and the level of blood pressure (BP) in women - two different approaches ...

## **LONGITUDINAL STUDY**

- identify a group of non-pregnant, pre-menopausal women of childbearing age who are not OC users and measure their BP (baseline value)
- rescreen the women after one year and find all the women who remained non-pregnant throughout the year and who have become OC users - this is the study population
- measure the BP of the women in the study population and compare the values to the baseline values (two groups of the same women, measuring BP at two different times)

## **CROSS-SECTIONAL STUDY**

- identify two groups of non-pregnant, pre-menopausal women of childbearing age, one of OC users and another that does not use OC, measure their BP
- compare the BP of the OC users and non-users (two groups of different women, measuring BP at the same time)

## **LONGITUDINAL STUDY**

- same group of people followed over time
- paired-sample design - each person within the group serves as his/her own control
- also possible to have measurements on two different people who have been matched to each other based on criteria such as age and sex
- the two samples are NOT independent

## **CROSS-SECTIONAL STUDY**

- different groups of people
- the two sample are INDEPENDENT

## **WHAT TYPE OF STUDY IS BETTER FOR OC AND BP**

- longitudinal helps to rule out other factors (confounders) that might also influence BP
- harder to rule out other factors in a cross-sectional study (statistical methods often used for adjustment)
- **IMPORTANT** to remember that different statistical methods are used to test for differences in **NON-INDEPENDENT** versus **INDEPENDENT** data

## INDEPENDENT SAMPLES

- proportions

proportions are based on data from two random samples that are independent (values in one sample are not related to values in the other), examples 10.1 and 10.2 in Rosner ...

compare the ovarian cancer rates between a group of "light OC users" (used OC for less than 5 years) and a group of "heavy OC users" (used OC for 5+ years)

Triola... for each sample, the NUMBER of successes and failures is at least 5 (same as saying  $NP \geq 5$  and  $NQ \geq 5$ )

Rosner... for each sample,  $NPQ \geq 5$  (normal approximation to the binomial)

- two-sample test for proportions (normal-theory)...

$$z = ((\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)) / \sqrt{(\bar{p}\bar{q} / n_1) + (\bar{p}\bar{q} / n_2)}$$

with a null hypothesis...

$$H_0: p_1 = p_2$$

reduces to...

$$z = (\hat{p}_1 - \hat{p}_2) / \sqrt{\bar{p}\bar{q}((1/n_1) + (1/n_2))}$$

where...

$$\bar{p} = (x_1 + x_2) / (n_1 + n_2)$$

$$\bar{q} = 1 - \bar{p}$$

Example from Rosner - what is the relationship between oral contraceptive (OC) use and heart disease in women... cross-sectional study ... example 10.6 ...

among women 40-44 years of age...

13 of 5,000 OC users have an MI over a 3-year period

7 of 10,000 non-OC users have an MI over a 3-year period

- given a claim...

OC use is related to occurrence of MI

- identify the null hypothesis...

the rate of MI occurrence is the same among OC and non-OC users

- identify the alternative hypothesis...

the rate of MI occurrence is not the same among OC and non-OC users

- express the null and alternative hypothesis in symbolic form...

$$H_0: p_1 = p_2$$

$$H_1: p_1 \neq p_2$$

- given a claim and sample data, calculate the value of the test statistic...

$$\bar{p} = (x_1 + x_2) / (n_1 + n_2) = (13 + 7) / (5000 + 10000) = 0.00133$$

$$\hat{p}_1 = x_1 / n_1 = 13 / 5000 = 0.0026$$

$$\hat{p}_2 = x_2 / n_2 = 7 / 10000 = 0.0007$$

$$z = (\hat{p}_1 - \hat{p}_2) / \sqrt{\bar{p}\bar{q}((1/n_1) + (1/n_2))}$$

$$z = (0.0026 - 0.0007) / \sqrt{(0.00133)(0.99867)((1/5000) + (1/10000))}$$

$$z = 0.0019 / 0.00063 = 3.02$$

- given a significance level, identify the critical value(s)...

normal distribution,  $\alpha = .05$ , 2-tail test

from table A-2 in Triola, critical value = 1.96

- given a value of the test statistic, identify the P-value...

in the normal distribution, what is the probability of obtaining the value of the test statistic 3.02

in table A-2 in Triola, P-value=0.0013

two-tail test, double the P-value, 0.0026

- state the conclusion of the hypothesis is simple, non-technical terms...

reject the null hypothesis

OC use is associated with MI occurrence

- identify the type I and type II errors that can be made when testing a given claim...

type I error determined by choice of  $\alpha$ , in this case, 0.05

type II error (calculated using web site), 0.19

- NOTE: Rosner uses a continuity correction in the numerator of the equation used to calculate the value of the test statistic...

$$\begin{aligned}
 & (\hat{p}_1 - \hat{p}_2) - ((1 / 2n_1) + (1 / 2n_2)) \\
 & (0.0026 - 0.0007) - ((1 / 2(5000)) + (1 / 2(10000))) \\
 & (0.0026 - 0.0007) - ((1 / 10000) + (1 / 20000)) = 0.00175
 \end{aligned}$$

applying the continuity correction leads to  $z=2.77$  (not 3.02)

- compute a confidence interval for the difference between the two proportions

the pooled estimate of  $p$  is NOT used in calculating the margin of error...

$$E = z_{\alpha/2} \sqrt{(\hat{p}_1 \hat{q}_1 / n_1) + (\hat{p}_2 \hat{q}_2 / n_2)}$$

$$E = 1.96 / \sqrt{((0.0026)(0.9974) / 5000) + ((0.0007)(0.9993) / 10000)}$$

$$E = 0.0015$$

confidence interval...

$$(\hat{p}_1 - \hat{p}_2) - E < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + E$$

$$(0.0026 - 0.0007) - 0.0015 < (p_1 - p_2) < (0.0026 - 0.0007) + 0.0015$$

$$0.0004 < (p_1 - p_2) < 0.0034$$

same conclusion as hypothesis test since 0 is outside the band

- compute a confidence band on both proportions and see if the bands overlap...

$$E_1 = z_{\alpha/2} \sqrt{\hat{p}_1 \hat{q}_1 / n_1}$$

$$E_1 = 1.96 \sqrt{(0.0026)(0.9974) / 5000} = 0.0014$$

$$\hat{p}_1 - E_1 < p_1 < \hat{p}_1 + E_1 = 0.0026 - 0.0014 < p_1 < 0.0026 + 0.0014$$

$$0.0012 < p_1 < 0.0040$$

$$E_2 = z_{\alpha/2} \sqrt{\hat{p}_2 \hat{q}_2 / n_2}$$

$$E_2 = 1.96 \sqrt{(0.0007)(0.9993) / 10000} = 0.0005$$

$$\hat{p}_2 - E_2 < p_2 < \hat{p}_2 + E_2 = 0.0007 - 0.0005 < p_2 < 0.0007 + 0.0005$$

$$0.0002 < p_2 < 0.0012$$

confidence bands "close" to overlap (actually do overlap with less rounding of the values) --- "...rejection of the null hypothesis by the method of overlap implies rejection by the standard method, whereas failure to reject...does NOT imply failure to reject by the standard method..." --- more conservative, less powerful method

## **NON-INDEPENDENT SAMPLES**

- from chapter 10 in Triola, McNemar's Test for matched-pairs
- previous method inappropriate for matched data
- test statistic is now  $\chi^2$ , not z
- 2x2 table used to display results
- used when subjects serve as their own controls or when different subjects are matched on one or more factors thought to be related to the outcome of interest
- analysis dependent on concordant and discordant pairs

Modified example from Rosner (example 10.25) ... 22 patients have their blood pressure evaluated by both a computerized device and the standard method (a person) ... the results look as follows with...  
H-hypertensive, N-normal BP

person	computer	standard	person	computer	standard
1	N	N	12	H	N
2	N	N	13	H	N
3	H	N	14	N	N
4	H	H	15	H	N
5	N	N	16	N	H
6	H	N	17	H	N
7	N	N	18	H	N
8	H	H	19	N	N
9	H	H	20	N	N
10	N	N	21	N	N
11	H	N	22	N	H

- question... are results different from the computer versus the person
- wrong approach... compute proportion of hypertensive found by each method...

$$p(\text{computer}) = 11/22 = 0.50$$

$$p(\text{standard}) = 5/22 = 0.23$$

and compare with method shown for independent samples  
(looks as if there are 44 subjects, 22 in each group)

- correct approach... McNemar's Test

- general...

	method 2	
method 1	yes	no
yes	concordant	discordant
no	discordant	concordant

- specific...

	standard	
computer	H	N
H	3	8
N	2	9

- requirements for using McNemar's Test

same issues as to sampling as before (random, representative)

sample data comprise matched-pairs

data are at a nominal level of measurement and each observation can be classified in two-ways (in the example... method used, BP category)

number of discordant pairs  $\geq 10$  (Rosner...  $\geq 20$ ), otherwise an exact method must be used

- test statistic is  $\chi^2$
- critical value located in right tail only with 1 degree of freedom

- claim... the two methods produce identical results
- null hypothesis...
  - $H_0: p_1 = p_2$
  - $H_1: p_1 \neq p_2$
- test statistic...
  - $\chi^2 = (|b - c| - 1)^2 / (b + c)$
  - $\chi^2 = (|8 - 2| - 1)^2 / (8 + 2) = 2.5$
- critical value... with  $\alpha = 0.05$ , 1 DF (table A-4), 3.841
- P-value... hard to determine from table in Triola  
use other tables (Rosner), or software  
from Statcrunch, P-value = 0.1138
- conclusion... fail to reject the null hypothesis  
no evidence to say the methods differ

- exact method (when number of discordant pairs  $< 10$ , but also works anytime)
- based on exact binomial probabilities

if  $n_a$  = number of discordant pairs in which the group A member has the event and the group B member does not

if  $n_d$  = total number of discordant pairs

binomial with  $p=0.5$ ,  $n=n_d$ ,  $k=n_a$

if  $n_a < n_d/2$ ,  $p = 2 \times$  probability of  $k$  events or less

if  $n_a > n_d/2$ ,  $p = 2 \times$  probability of  $k$  events or more

if  $n_a = n_d/2$ ,  $p = 1$

- change the previous problem...

	standard	
computer	H	N
H	3	7
N	1	9

- number of discordant pairs = 10
- $n_a = 7, n_d = 8, n_a > n_d/2 (7 > 4)$   
 binomial with  $p=0.5, n=8, k=7$   
 $p$  of  $k$  or more events = 0.03515  
 exact  $p$ -value =  $2 \times 0.03515 = 0.0703$

- previous problem

$n_a = 8, n_d = 10, n_a > n_2/2 (8 > 5)$

binomial with  $p=0.5, n=10, k=8$

$p$  of  $k$  or more events = 0.05468

exact  $p$ -value =  $2 \times 0.05468 = 0.1093$

using SAS, same answer... exact = 0.1094

**Table of COMPUTER by STANDARD**

COMPUTER	STANDARD		Total
Frequency	H	N	
H	3	8	11
N	2	9	11
Total	5	17	22

**Statistics for Table of COMPUTER by STANDARD**

**McNemar's Test**

Statistic (S)	3.6000
DF	1
Asymptotic Pr > S	0.0578
Exact Pr >= S	0.1094

- another type of matching, not same subjects...

example from Rosner (example 10.21) ... compare two different regimens of chemotherapy for breast cancer after mastectomy... matched study using two groups of women where pairs of women are matched on age and clinical condition... random member of each pair gets treatment A, other gets treatment B... outcome measure is survival at the end of 5 years of follow up

621 pairs (1242 women)

526 ( $p=0.847$ ) in group A survive

515 ( $p=0.829$ ) in group B survive

- table with matched pairs...

	treatment B	
treatment A	survived	died
survived	510	16
died	5	90

- claim... the two methods produce identical results
- null hypothesis...  $H_0: p_1 = p_2$   
 $H_1: p_1 \neq p_2$
- test statistic...  $\chi^2 = (|b - c| - 1)^2 / (b + c)$   
 $\chi^2 = (|16 - 5| - 1)^2 / (16 + 5) = 4.76$
- critical value... with  $\alpha = 0.05$ , 1 DF (table A-4), 3.841
- P-value... 4.76 between 3.841 ( $p = 0.05$ ) and 5.024 ( $p = 0.025$ ) in Triola table A-4, or from Statcrunch, P-value = 0.029
- conclusion... reject the null hypothesis  
survival different for different treatments

## **MATCHED PAIRS**

- continuous data
- test statistic is t
- requirements

same issues as to sampling as before (random, representative)

sample data comprise matched pairs

either or both conditions: number of matched pairs  $> 30$ ;  
distribution of differences between matched pairs is  
approximately normally distributed

Example from Rosner... matched pairs, same person...

blood pressure measurements taken on both the left and right arms of 10 subjects... are the measurements comparable

subject	left	right	difference
1	130	126	4
2	120	124	-4
3	135	127	8
4	100	95	5
5	98	102	-4
6	110	109	1
7	123	124	-1
8	136	132	4
9	140	137	3
10	155	156	-1

- claim... measurements on the two arms produce identical results
- null hypothesis...  $H_0: \mu_d = 0$   
 $H_1: \mu_d \neq 0$
- test statistic...  $t = \bar{d} / (s_d / \sqrt{n})$   
 $t = 1.50 / (3.979 / \sqrt{10}) = 1.19$
- critical value... with  $\alpha = 0.05$ , 9 DF, two-tails (table A-3), 2.262
- P-value... hard to determine from table in Triola  
use other tables (Rosner), or software  
from Statcrunch, P-value = 0.264
- conclusion... fail to reject the null hypothesis  
no evidence to say the BP differs from one  
arm to the other

- confidence interval for matched pairs

$$E = t_{\alpha/2}(s_d / \sqrt{n})$$

$$E = 2.262(3.979 / \sqrt{10}) = 2.85$$

$$\bar{d} - E < \mu_d < \bar{d} + E$$

$$1.50 - 2.85 < \mu_d < 1.50 + 2.85$$

$$-1.35 < \mu_d < 4.35$$

same conclusion as hypothesis test since zero (no difference) is within the confidence band

Example from Rosner...  
 matched pairs, different  
 persons...

30 pregnant women, 15  
 matched pairs (age and  
 weight as matching criteria,  
 women are within 5 pounds  
 of each other and within 2  
 years of age), member of  
 each pair randomly  
 assigned to a drug or  
 placebo to see if the drug is  
 effective in reducing  
 premature births (outcome  
 measure is birth weight, in  
 pounds)

Subject	Treatment	Control	Difference
1	6.9	6.4	0.5
2	7.6	6.7	0.9
3	7.3	5.4	1.9
4	7.6	8.2	-0.6
5	6.8	5.3	1.5
6	7.2	6.6	0.6
7	8.0	5.8	2.2
8	5.5	5.7	-0.2
9	5.8	6.2	-0.4
10	7.3	7.1	0.2
11	8.2	7.0	1.2
12	6.8	6.9	-0.1
13	6.8	5.6	0.8
14	5.7	4.2	1.5
15	8.6	5.8	2.8

- claim... drug treatment helps to prevent prematurity
- null hypothesis...
  - $H_0: \mu_d = 0$
  - $H_1: \mu_d > 0$
- test statistic...
  - $t = \bar{d} / (s_d / \sqrt{n})$
  - $t = 0.820 / (0.887 / \sqrt{15}) = 3.58$
- critical value... with  $\alpha=0.05$ , 14 DF, one-tail (table A-3), 1.761
- P-value... hard to determine from table in Triola use other tables (Rosner), or software from Statcrunch, P-value = 0.002
- conclusion... reject the null hypothesis  
drug treatment helps to prevent prematurity

- confidence interval for matched pairs (one-sided)

$$E = t_{\alpha/2}(s_d / \sqrt{n})$$

$$E = 1.761(0.887 / \sqrt{15}) = 0.403$$

$$\bar{d} - E < \mu_d$$

$$0.820 - 0.403 < \mu_d$$

$$0.417 < \mu_d$$

same conclusion as hypothesis test since zero (no difference) not within the lower bound of the one-sided confidence band

## TWO MEANS - INDEPENDENT SAMPLES

- continuous data
- test statistic is  $t$
- requirements

same issues as to sampling as before (random, representative)

two samples are independent

either or both conditions: in both sample, sample size  $> 30$ ;  
both samples come from populations having normal distributions  
(test is robust in that departures from normality requirement is not strict as long as there are no outliers and that data are not extremely far from being normally distributed)

## ISSUES

- how to compute the standard error of the difference between two means
- how to compute the degrees of freedom to use when looking up the critical values and when computing P-values

- compare the variances from the two samples

if equal... compute a pooled estimate of the standard error...

$$se = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} \quad \text{where... } sp^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

or...

$$se = sp \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad \text{where... } sp = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

and degrees of freedom...

$$df = n_1 + n_2 - 2$$

if not equal... Satterthwaite's Method...

$$se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

and degrees of freedom...

$$df = \frac{((s_1^2 / n_1) + (s_2^2 / n_2))^2}{((s_1^2 / n_1)^2 / (n_1 - 1)) + ((s_2^2 / n_2)^2 / (n_2 - 1))}$$

(if not a whole number, round down to the nearest integer)

Triola... degrees of freedom...

$$df = \text{smaller of } (n_1 - 1) \text{ and } (n_2 - 1)$$

## **TWO VARIANCES - INDEPENDENT SAMPLES**

- test statistic is F
- requirements

same issues as to sampling as before (random, representative)

two samples are independent

two populations are normally distributed (important since methods used to compare variances are extremely sensitive to departures from normality)

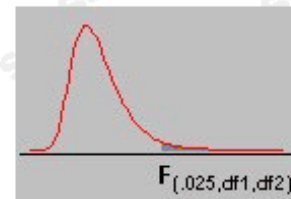
Example from Triola... summary statistics on male and female students (in inches)...

*males:*  $n = 16, \bar{x} = 68.4, s = 2.54$

*females:*  $n = 12, \bar{x} = 63.2, s = 2.39$

- claim... the variances in the two groups are equal
- null hypothesis...
  - $H_0: \sigma_1^2 = \sigma_2^2$
  - $H_1: \sigma_1^2 \neq \sigma_2^2$
- test statistic...  $F = s_1^2 / s_2^2 = 2.54^2 / 2.39^2 = 1.1295$
- critical value... with  $\alpha = 0.025$  (in right tail), 15 DF in numerator, 11 DF in denominator, table A-5 in Triola, 3.3299

F Table for alpha=.025 .



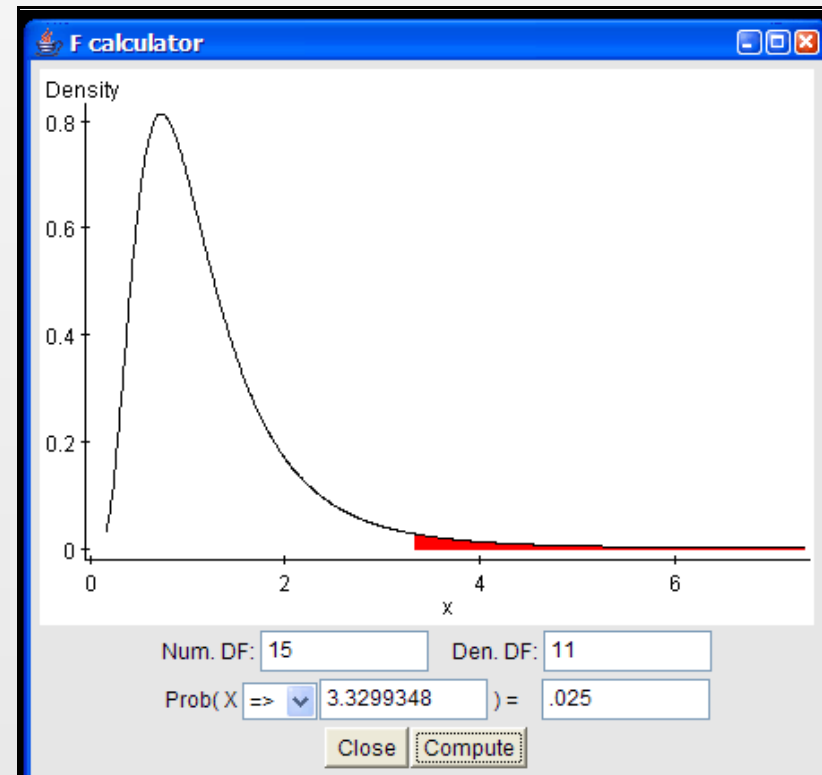
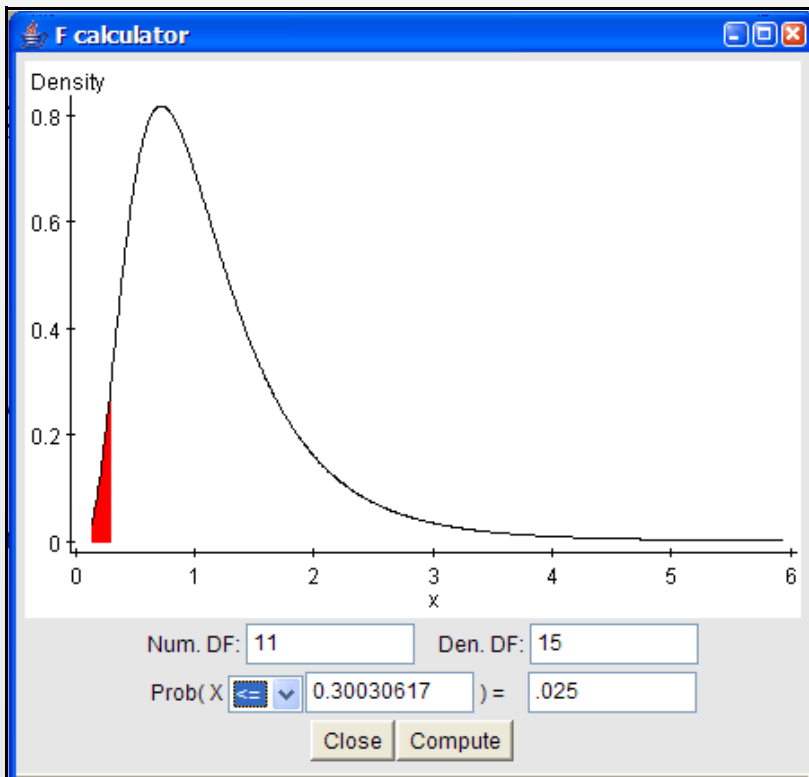
df2/df1	1	2	3	4	5	6	7	8	9	10	12	15
1	647.7890	799.5000	864.1630	899.5833	921.8479	937.1111	948.2169	956.6562	963.2846	968.6274	976.7079	984.8668
2	38.5063	39.0000	39.1655	39.2484	39.2982	39.3315	39.3552	39.3730	39.3869	39.3980	39.4146	39.4313
3	17.4434	16.0441	15.4392	15.1010	14.8848	14.7347	14.6244	14.5399	14.4731	14.4189	14.3366	14.2527
4	12.2179	10.6491	9.9792	9.6045	9.3645	9.1973	9.0741	8.9796	8.9047	8.8439	8.7512	8.6565
5	10.0070	8.4336	7.7636	7.3879	7.1464	6.9777	6.8531	6.7572	6.6811	6.6192	6.5245	6.4277
6	8.8131	7.2599	6.5988	6.2272	5.9876	5.8198	5.6955	5.5996	5.5234	5.4613	5.3662	5.2687
7	8.0727	6.5415	5.8898	5.5226	5.2852	5.1186	4.9949	4.8993	4.8232	4.7611	4.6658	4.5678
8	7.5709	6.0595	5.4160	5.0526	4.8173	4.6517	4.5286	4.4333	4.3572	4.2951	4.1997	4.1012
9	7.2093	5.7147	5.0781	4.7181	4.4844	4.3197	4.1970	4.1020	4.0260	3.9639	3.8682	3.7694
10	6.9367	5.4564	4.8256	4.4683	4.2361	4.0721	3.9498	3.8549	3.7790	3.7168	3.6209	3.5217
11	6.7241	5.2559	4.6300	4.2751	4.0440	3.8807	3.7586	3.6638	3.5879	3.5257	3.4296	3.3299
12	6.5538	5.0959	4.4742	4.1212	3.8911	3.7283	3.6065	3.5118	3.4358	3.3736	3.2773	3.1772

- reverse variances in calculation of test statistic...

$$F = \sigma_2^2 / \sigma_1^2 = 2.39^2 / 2.54^2 = 0.8854$$

look in the left-tail, critical value is now 0.3003

NOTE: after reversing the degrees of freedom,  
left-tail critical value = 1 / right-tail critical value

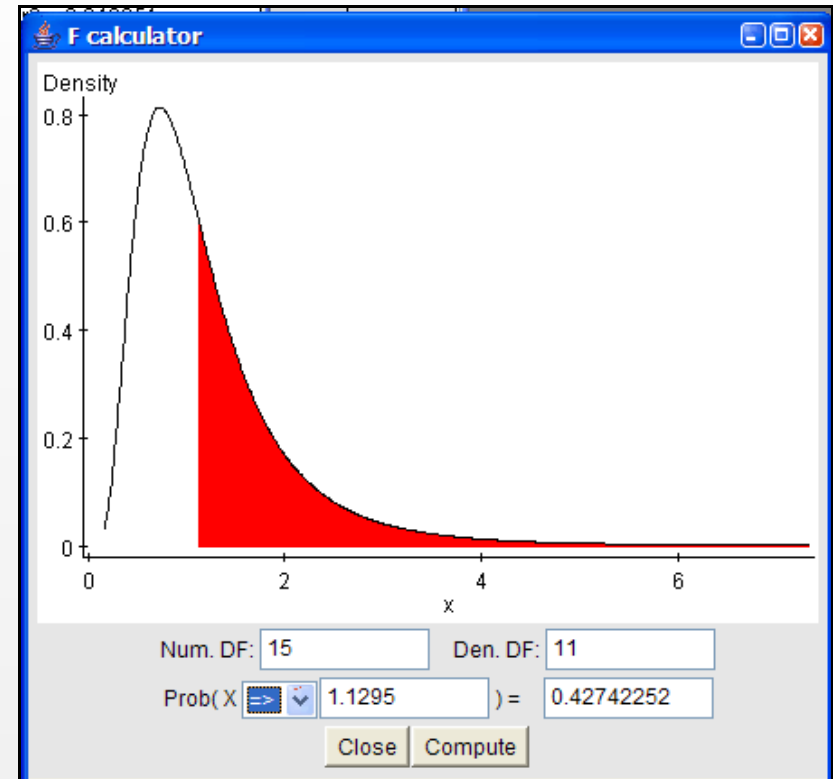


- P-value...

hard to determine from any F table, use Statcrunch...

P-value = 0.8548

(2x value from Statcrunch since this is two-tail test)



- conclusion... fail to reject the null hypothesis  
no evidence to say the variances differ
- Rosner... "Be cautious about using this test with nonnormally distributed samples."

- confidence interval on the ratio of two variances...

$$\left(\frac{s_1^2}{s_2^2}\right)\left(\frac{1}{F_R}\right) < \frac{\sigma_1^2}{\sigma_2^2} < \left(\frac{s_1^2}{s_2^2}\right)\left(\frac{1}{F_L}\right)$$

$$\left(\frac{2.54^2}{2.39^2}\right)\left(\frac{1}{3.3299}\right) < \frac{\sigma_1^2}{\sigma_2^2} < \left(\frac{2.54^2}{2.39^2}\right)\left(\frac{1}{0.3325}\right)$$

$$0.3392 < \frac{\sigma_1^2}{\sigma_2^2} < 3.3969$$

- same conclusion... no evidence to say the variances differ  
confidence interval includes 1

## WHAT SHOULD YOU DO

- Triola suggests...

ignore the preliminary variance test

do not pool the sample variances

use the smaller sample size in the two groups to find DF

- others...

perform a preliminary test to see if variances are equal

if equal, use a pooled estimate of the variance and use...

$$DF = n_1 + n_2 - 2$$

if unequal, do not pool the sample variances and use Satterthwaite's method to compute DF

- for quiz problems, you should...

perform a preliminary test to see if variances are equal

if equal, use a pooled estimate of the variance and use...

$$DF = n_1 + n_2 - 2$$

if unequal, do not pool the sample variances and use the smaller sample size in the two groups to find DF (don't bother to use Satterthwaite's method)

Example from Triola... effectiveness of paroxetine for treating bipolar depression... scores on a depression scale...

*placebo group:  $n = 43, \bar{x} = 21.57, s = 3.87$*

*treatment group:  $n = 33, \bar{x} = 20.38, s = 3.91$*

- claim... there is no difference between the placebo and treatment groups

- null hypothesis...  
 $H_0: \mu_1 = \mu_2$   
 $H_1: \mu_1 \neq \mu_2$

- compare variances...  $F = 3.91^2 / 3.87^2 = 1.0208$

critical value with  $\alpha=0.05$ , 32 DF in numerator, 42 DF in denominator, 1.9078 (no evidence to reject null hypothesis of equal variances)

- pool variances... 
$$sp = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$sp = \sqrt{\frac{(42)3.87^2 + (32)3.91^2}{74}} = 3.8873$$

- test statistic... 
$$t = (\bar{x}_1 - \bar{x}_2) / sp\sqrt{(1/n_1) + (1/n_2)}$$
  

$$t = (21.57 - 20.38) / 3.89\sqrt{(1/43) + (1/33)} = 1.323$$

NOTE: test statistic in text without using pooled variances, t=1.321

- critical value... with  $\alpha=0.05$ , 74 DF, two-tail from Statcrunch, 1.993

- P-value... hard to determine from table in Triola from Statcrunch, P-value = 0.19

- conclusion... no evidence to reject the null hypothesis

- confidence interval on the difference between two means...

$$E = t_{\alpha/2} (sp \sqrt{\frac{1}{n1} + \frac{1}{n2}})$$

$$E = 1.993(3.89 \sqrt{\frac{1}{43} + \frac{1}{33}})$$

$$E = 1.79$$

$$(\bar{x}_1 - \bar{x}_2) - E < (\mu_1 - \mu_2) < (\bar{x}_1 - \bar{x}_2) + E$$

$$(21.57 - 20.38) - 1.79 < (\mu_1 - \mu_2) < (21.57 - 20.38) + 1.79$$

$$- 0.603 < (\mu_1 - \mu_2) < 2.983$$

- same conclusion... no evidence to reject the null hypothesis  
confidence interval includes 0

NOTE: confidence interval without using paired variances...

$$- 0.607 < (\mu_1 - \mu_2) < 2.987$$

## SAMPLE SIZE AND POWER

- sample size for comparing means of two normally distributed samples of equal size using a two-sided significance test...

$$n = \frac{(\sigma_1^2 + \sigma_2^2)(z_{1-\alpha/2} + z_{1-\beta})^2}{\Delta^2}$$

where...  $\Delta = |\mu_1 - \mu_2|$  (difference worth detecting)

how does sample size vary with... variability,  $\alpha$ ,  $\beta$ ,  $\Delta$

- power for comparing means of two normally distributed samples of equal size using a two-sided significance test...

$$power = \Phi \left[ -z_{1-\alpha/2} + \frac{\Delta}{\sqrt{(\sigma_1^2 / n_1) + (\sigma_2^2 / n_2)}} \right]$$

where...  $\Delta = |\mu_1 - \mu_2|$  (difference worth detecting)

how does power vary with... variability,  $\alpha$ ,  $n$ ,  $\Delta$

- sample size for longitudinal studies comparing mean change in two normally distributed samples with two time points using a two-sided significance test...

$$n = \frac{2\sigma_d^2(z_{1-\alpha/2} + z_{1-\beta})^2}{\Delta^2}$$

where...  $\Delta = |\mu_1 - \mu_2|$  (difference worth detecting)

$\sigma_d^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$  ( $\rho$  is the correlation coefficient between baseline and followup values)

how does sample size vary with... variability,  $\alpha$ ,  $\beta$ ,  $\rho$ ,  $\Delta$

- power of a longitudinal studies comparing mean change in two normally distributed samples with two time points using a two-sided significance test...

$$power = \Phi \left[ -z_{1-\alpha/2} + \frac{\sqrt{n}\Delta}{\sigma_d \sqrt{2}} \right]$$

where...  $\Delta = |\mu_1 - \mu_2|$  (difference worth detecting)  
 $\sigma_d^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$  ( $\rho$  is the correlation coefficient between baseline and followup values)

how does power vary with... variability,  $\alpha$ ,  $n$ ,  $\rho$ ,  $\Delta$