Chapter 15
Synthesis of Findings

Much of the research conducted in education, psychology, business, and other disciplines has involved single experiments or studies that rarely provide definitive answers to research questions. We have learned that researchers seldom replicate their research studies and instead use cross-validation, jackknife, or bootstrap methods to estimate the stability and accuracy of a sample statistic as an estimate of a population parameter. The world around us is understood better when we discover underlying patterns, trends, and principles, which can result from an accumulation of knowledge gained from several studies on a topic of interest. Consequently, a review of the research literature is invaluable in summarizing and understanding the current state of knowledge about a topic. Rather than rely on subjective judgments or interpretations of the research literature, meta-analysis techniques provide a quantitative objective assessment of the study results.

Meta-Analysis

Meta-analysis is the application of statistical procedures to the empirical findings in research studies for the purpose of summarizing and concluding whether the findings in the studies overall were significant. The meta-analytic approach therefore provides a method to synthesize research findings from several individual studies. It makes sense that if several hundred studies had researched socioeconomic status and achievement in school, that a summarization of the findings in these individual studies would aid our understanding of this relationship. From a scholarly point of view, we might ask, "After 20 years of studying the efficacy of psychotherapy, what have we learned?" At some point in time, it becomes frugal to assess what we have learned from the research that was conducted. Further research in an area may be unproductive, unscientific, or take a wrong direction, if we don't stop from time to time and assess what we have learned. If we view science as an objective activity, then it becomes critical to establish an objective method to integrate

R. Schumacker and S. Tomek, *Understanding Statistics Using R*,
DOI 10.1007/978-1-4614-6227-9_15,

and synthesize similar research studies. Meta-analysis is therefore an objective analysis of the statistical analyses from several individual studies for the purpose of integrating and summarizing the findings.

Sir Ronald A. Fisher in 1932, Karl Pearson in 1933, and Egon S. Pearson in 1938 all independently addressed the issue of statistically summarizing the results of research studies. In 1952, Mordecai H. Gordon, Edward H. Loveland, and Edward E. Cureton produced a chi-square table for use in combining the probability values from independent research studies. In 1953, Lyle V. Jones and Donald W. Fiske further clarified their approaches for testing the significance of results from a set of combined studies. They further demonstrated Fisher's approach of taking a natural logarithm of a p-value to calculate a summary chi-square value. Gene Glass in 1976 is credited with using the term *meta-analysis* to describe the statistical analysis of a collection of analysis results from several individual studies. He provided guidelines for converting various statistics into a common metric. Jacob Cohen in 1965 and again in 1977 provided measures of effect size for many common statistical tests. An effect size measure indexes the degree of departure from the null hypothesis in standard units. Examples of these various approaches to combining results from several research studies will be presented in this chapter.

A Comparison of Fisher and Gordon Chi-Square Approaches

The Fisher approach to combining p-values from several independent research studies was accomplished by using a natural logarithmic transformation of the p-value (The natural log or log base *e* is equal to 2.7182818). The sum of the log base *e* values times −2 resulted in a chi-square value: $\chi^2 = -2\ \Sigma\ (\log_e p)$, with degrees of freedom equal to 2n, where n = the number of studies. The following example helps to illustrate Fisher's approach and the tabled chi-square values provided by Gordon, Loveland, and Cureton.

Research Study	p	Fisher $\log_e p$	Gordon et al. Tabled χ^2
1	.05	−2.996	5.991
2	.01	−4.605	9.210
3	.04	−3.219	6.438
Total		−10.820	21.639

Fisher's approach required multiplying −2 times the sum of the natural log values to calculate the chi-square value: −2 (−10.820) = 21.639! Gordon et al. produced a chi-square table with the chi-square values for various p-values, thus making the task easier!

Converting Various Statistics to a Common Metric

Gene Glass discovered that various test statistics reported in the research literature could be converted to the Pearson Product Moment Correlation Coefficient or *r*, e.g., the *t*, chi-square, and *F*-values. This provided a common metric to compare the various statistical values reported in research studies. The formula for transforming *r* to each statistic is presented below.

t-test

$$r = \sqrt{\frac{t^2}{t^2 + df}}$$

F-test

$$r = \sqrt{\frac{F}{F + df_{error}}}$$

Chi-square

$$r = \sqrt{\frac{\chi^2}{n}}$$

Converting Various Statistics to Effect Size Measures

Jacob Cohen expanded Gene Glass's idea to include a formula that would use an **effect size** measure (*d*) in the computation of the correlation statistic. The effect size formula is:

Effect-Size

$$r = \frac{d}{\sqrt{d^2 + 4}}$$

The formula for transformation to an effect size measure *d* for each statistic is presented below.

t-test

$$d = \frac{2t}{\sqrt{df}}$$

F-test

$$d = \frac{2\sqrt{F}}{\sqrt{df_{error}}}$$

r

$$d = \frac{2r}{\sqrt{1 - r^2}}$$

Comparison and Interpretation of Effect Size Measures

The concept behind an effect size measure, *d*, was to determine the amount of departure from the null hypothesis in standard units. Consequently, an effect size measure in an experimental-control group study was determined by the following formula:

$$d = \frac{\overline{Y}_{EXP} - \overline{Y}_{CTRL}}{S_{CTRL}}$$

If $d = .50$, then the experimental group scored one-half standard deviations higher than the control group. If the population standard deviation is known, it would be used instead of the control group sample standard deviation estimate. Another alternative was to use the pooled estimate of the standard deviation from both groups in the denominator.

Not all research studies however used an experimental-control group design, so the development of other effect size formulae were very important in being able to compare the results of various studies. A comparison of *r* and *d* effect size measures for a set of studies should help to better understand how the two methods are computed and interpreted. A comparison of *r* and *d* is listed below for four studies, each with a different statistic reported.

					Effect size measures	
Study	Statistic	N	df	p (one-tail)	r	d
1	$t = 2.70$	42	40	.005	.3926	.8538
2	$F = 4.24$	27	1,25*	.025	.3808	.8236
3	$\chi^2 = 3.84$	100	1	.05	.1959	.3995
4	$r = .492$	22	20	.01	.4920	1.1302

**In the meta-analysis program only the degree of freedom error is input. This is the second degree of freedom listed in the F ANOVA table*

The calculations are straightforward using the formula for the *r* and *d* effect size measures. For example, given that $t = 2.70$, the *r* effect size is computed as:

$$r = \sqrt{\frac{t^2}{t^2 + df}} = \sqrt{\frac{(2.7)^2}{(2.7)^2 + 40}} = \sqrt{\frac{7.29}{47.29}} = \sqrt{.1542} = .3926$$

The *d* effect size is computed as:

$$d = \frac{2t}{\sqrt{df}} = \frac{2(2.7)}{\sqrt{40}} = \frac{5.4}{6.3245} = .8538$$

The calculation of the *r* effect size measure for the chi-square value is straightforward; however, the resulting *r* effect size value should be used in the formula for computing the corresponding *d* effect size measure. Given chi-square = 3.84 with a sample size of 100, the *r* effect size measure is .1959. This *r* effect size value is used in the *d* effect size formula to obtain the value of .3995.

Since the various statistics from the different research studies are now on a common metric, they can be compared. Study 4 had the highest *r* effect size measure, followed by studies 1, 2, and 3, respectively. The corresponding *d* effect size measures also indicate the same order, but help our interpretation by indicating how much the dependent variable (Y) changed for unit change in the independent variable (X). Notice this interpretation is directly related to the use of the correlation coefficient as an effect size measure. If we had used the experimental-group effect size measure, our interpretation would be how much the experimental group increased on average over the control group with regard to the dependent variable.

The null hypothesis always implies that the effect size is zero. If the alternative hypothesis is accepted, then the effect size departs from zero and leads to a standardized interpretation. Jacob Cohen in 1977 offered basic guidelines for interpreting the magnitude of the *d* effect size measure (d = .2 = small, d = .5 = medium, and d = .8 = large) and *r* effect size measure (r = .10 = small, r = .30 = medium, r = .50 = large); however, any interpretation of an effect size measure is relative. Knowledge of the professional discipline and the distribution of effect size measures in similar research studies provide reference points for interpretation. Sometimes, no such distributions of effect size estimates exist, so no standard reference point is available. In this instance, the computation of effect size estimates in several hundred studies can provide the necessary distribution and reference point. Summary statistics in the stem and leaf procedure will facilitate the interpretation of the distribution of *d* and *r* effect size measures. Given the standard deviation unit interpretation, it is important to determine whether a one-half, one-third, or one-quarter standard deviation improvement is due to some type of intervention that implies a meaningful effect size for the interpretation of the research study outcome.

A basic approach in conducting a meta-analysis across several related research studies is to combine the results and determine if there was an *overall* significant effect. The use of p-values from research studies (see Fisher, Jones, and Fiske, and Gordon et al.) readily lends itself to a summary chi-square value. The chi-square value indicates the significance of the combined research study results. The meta-analysis formula for combining the p-values from individual studies is $\chi^2 = -2\ \Sigma\ (\log_e p)$. The combined chi-square value reported earlier was 21.6396. This chi-square value is tested for significance using 2 n degrees of freedom (df = 6) to determine the overall effect across the different research studies. The tabled chi-square value at $p < 0.01$ for 6 degrees of freedom is 16.812. The combined chi-square of 21.6396 exceeds this tabled value and therefore indicates an overall significant effect across the three research studies. In the case of using an *r* effect size measure, the correlation coefficients converted from the various statistics are simply averaged, using the formula:

$$\overline{r} = \frac{\Sigma r}{n}$$

In the previous example, the overall r effect size is:

$$\overline{r} = \frac{\Sigma r}{n} = \frac{(.3926 + .3808 + .1959 + .4920)}{4} = \frac{1.4613}{4} = .3653$$

In the case of using a d effect size measure, the individual values are also averaged:

$$\overline{d} = \frac{\Sigma d}{n} = \frac{(.8538 + .8236 + .3995 + 1.1302)}{4} = \frac{3.2071}{4} = .8018$$

Sample Size Considerations in Meta-Analysis

An important concern in combining studies using meta-analytic techniques is the influence that different sample sizes may have on the overall interpretation. Some studies may be based on small sample sizes whereas others may be based on larger sample sizes. Since the correlation coefficient is not a function of sample size, procedures were developed to take into account the different sample sizes from the research studies when averaging the d effect size estimates. L. Hedges, as well as, R. Rosenthal and D. Rubin separately developed a formula in 1982 for calculating an *unbiased estimate* of the average d effect size. The formula was:

$$\overline{d} = \frac{\Sigma wd}{\Sigma w}$$

with w calculated as (N = total sample size):

$$w = \frac{2N}{8 + d^2}$$

The results of applying this unbiased, weighted approach to the previous example are listed below:

Study	N	d	w
1	42	.8538	9.623
2	27	.8236	6.222
3	100	.3995	24.511
4	22	1.1302	4.743

The calculation of the unbiased, weighted average effect estimate is:

$$\bar{d} = \frac{\Sigma wd}{\Sigma w}$$

$$= \frac{(9.623)(.8538) + (6.222)(.8236) + (24.5111)(.3995) + (4.743)(1.1302)}{9.623 + 6.222 + 24.511 + 4.743}$$

$$= \frac{28.493}{45.099} = .6318$$

The *unbiased* weighted average *d* effect size measure of .6318 is then compared to the *biased* weighted average *d* effect size measure of .8018 reported earlier. The amount of bias, or .8018 – .6318, is .17, which is due to the research studies having very different sample sizes. In practice, we would report the unbiased, weighted *d* effect size.

Meta-analysis is an objective quantitative method for combining the results of several independent research studies. One approach to combining research findings uses the log of p-values. The overall significant effect using p-values is indicated by a chi-square value. The experimental-control group effect size estimate is determined by subtracting the two group means and dividing by the standard deviation of the control group. The effect size indicates the departure from the null hypothesis in standard units. Other approaches to combining research findings use *r* and *d* effect size estimates, which involve the transformation of several common statistics. The overall significant effect from several studies using transformed statistics is obtained by averaging either the *r* or *d* effect size measures. Meta-analysis compares the relative importance of findings in several research studies by interpreting effect size measures, which are on a common metric. The overall *d* effect size measure can be weighted by sample size to compute an unbiased, average *d* effect size measure.

META-ANALYSIS R Programs

The **Meta-Analysis** program enters the p-values for each research study. Next, the chi-square values are computed given the p-values. The Fisher ln(p) and the Gordon et.al. values are then printed. The overall chi-square, degrees of freedom, and p-value are printed. A statistically significant chi-square indicates that the combined studies overall had a significant effect.

The **Effect Size** program enters for each study the following information: sample size, degree of freedom, p-value, and statistic. Each study can have a different statistic. The different statistics are converted to r and d effect size values. Next, the sample weight value is computed and used to calculate the unbiased effect size measure and sample size bias. The entire set of inputted values along with the r, d, and w values are printed. The average r and d effect size measures for the set of studies is printed along with the unbiased effect size and sample size bias.

Meta-Analysis Program Output

```
Fisher ln(p) versus Gordon et. al Chi-square

 Study      p     Fisher ln(p)     Gordon et al
     1  0.05           -2.996             5.991
     2  0.01           -4.605             9.210
     3  0.04           -3.219             6.438

 Chi-square = 21.64 df = 6 p = 0.00141
```

Effect Size Program Output

```
Effect Size r and d

 Type    Statistic  N   df    p      r      d      w
 t        2.7       42  40    0.005  0.393  0.855  9.621
 F        4.24      27  25    0.025  0.381  0.824  6.222
 Chisq    3.84      100 1     0.05   0.196  0.4    24.51
 r        0.492     22  20    0.01   0.492  1.13   4.743

 Effect size(r) = 0.366 Effect size(d) = 0.802

 Unbiased effect size(d) = 0.632

 Sample Size bias = 0.17
```

Note: For F values, **only** use the degrees of freedom error or denominator degree of freedom in the dialog box.

Meta-Analysis Exercises

1. Run the Meta-Analysis program for the p-values from the research studies below.
   Record the corresponding Fisher log base *e* values, total, and overall chi-square, df, and *p*.

Research Study	p	Fisher $\log_e p$
1	.05	
2	.001	
3	.20	
4	.01	
5	.025	
Total		

   Chi-Square=______, df=_______, p=_____

   a. Compute the chi-square value (−2 times the sum of $\log_e p$ values) with 2n degrees of freedom. Note: n = number of studies.

$\chi^2 = -2\ \Sigma\ (\log_e p)$: ________________

df = 2n: _________

b. Compare the chi-square above to the tabled chi-square value in Table A4 in Appendix (use .05 level of significance). What would you conclude?

____

____

____

2. Run the Effect Size program to compute *r* and *d* effect size estimates for the statistical values reported in the research studies below. Record the values in the table.

					Effect size measures	
Study	Statistic	N	df	p (one-tail)	r	d
1	t = 2.617	122	120	.005		
2	F = 4.000	62	60	.025		
3	$\chi^2 = 6.635$	50	1	.01		
4	r = .296	32	30	.05		

a. What is the overall average *r* effect size? ______________________

b. What is the overall average *d* effect size? ______________________

c. What would you conclude about the research findings from these results?

____

____

____

3. Run the META-ANALYSIS program again using the p-values from Exercise 2. Record the chi-square value and degrees of freedom. Select a tabled chi-square value for $p < .05$.

$\chi^2 = -2\ \Sigma\ (\log_e p)$: ___________ Tabled χ^2 = ___________

df = 2n: _________ df = 2n: ___________

a. What would you conclude about the research findings using the p-value approach?

____

____

b. Compare the chi-square, r, and d effect size results. What would you conclude?

____

____

4. Run the Effect Size program using the sample sizes from the research studies in Exercise 2. Record the d effect size measures computed in Exercise 2 and the weight values in the table below:

Study	N	d	w	wd
1	122			
2	62			
3	50			
4	32			
			Σw = ____	Σwd = ____

a. Compute the unbiased, average d effect size using the formula:

$$\bar{d} = \frac{\Sigma wd}{\Sigma w}$$

Unbiased Effect Size = ______________________

b. Compare the bias effect size (d) in Exercise 2 with the unbiased effect size ($\bar{d}$) above. How much bias in the overall effect size is due to the research studies having different sample sizes? Note: Overall Bias = (Bias Effect Size – Unbiased Effect Size) = $d - \bar{d}$.

____

____

Statistical Versus Practical Significance

Statistical tests for research questions involve tests of null hypotheses for different types of statistics. The statistical tests were the chi-square, z-test, t-test, analysis of variance, correlation, and linear regression. The outcomes of the statistical tests were to either retain the null hypothesis or reject the null hypothesis in favor of an alternative hypothesis based on the significance of the statistic computed. TYPE I and TYPE II errors were illustrated to better understand the nature of falsely rejecting the null hypothesis or falsely retaining the null hypothesis at a given level of significance for the sample statistic. The level of significance or p-value that we choose, i.e., .05 or .01, to test our null hypothesis has come under scrutiny due to the nature of statistical significance testing.

Researchers have criticized significance testing because it can be manipulated to achieve the desired outcome, namely, a significant finding. This can be illustrated by presenting different research outcomes based on only changing the p-value selected for the research study. The research study involves fifth-grade boys and girls who took the Texas Assessment of Academic Skills (TAAS) test. The study was interested in testing whether fifth-grade boys on average scored statistically significantly higher than girls on the TAAS test (a directional or one-tailed test of the null hypothesis). The researcher took a random sample of 31 fifth-grade boys and 31 fifth-grade girls and gave them the TAAS test under standard administration conditions. An independent *t*-test was selected to test for mean differences between the groups in the population at the .01 level of significance with 60 degrees of freedom (df = N − 2). The resultant sample values were:

Group	N	Mean	Standard deviation	t
Boys	31	85	10	1.968
Girls	31	80	10	

The researcher computed the t-value as follows:

$$t = \frac{85-80}{\sqrt{\frac{30(100)+30(100)}{31+31-2}\left(\frac{1}{31}+\frac{1}{31}\right)}} = \frac{5}{2.54} = 1.968$$

The tabled t-value that was selected for determining the research study outcome (based on a directional, one-tailed test, with 60 degrees of freedom at the .01 level of significance) was t = 2.39. Since the computed t = 1.968 was not greater than the tabled t-value of 2.66 at the .01 level of significance, the researcher would *retain* the null hypothesis. However, if the researcher had selected a .05 level of significance, the tabled t-value would equal 1.67, and the researcher would *reject* the null hypothesis in favor of the alternative hypothesis. The two possible outcomes in the research study are due solely to the different levels of significance a researcher could choose for the statistical test. This points out why significance testing has been criticized, namely the researcher can have statistically significant research findings by simply changing the p-value.

Researchers could also manipulate whether statistically significant results are obtained from a research study by using a *one-tailed test* rather than a *two-tailed test*. In the previous example, a two-tailed test of significance would have resulted in a tabled t = 2.66 at the .01 level of significance or a tabled t = 2.00 at the .05 level of significance. If the researcher had chosen a two-tailed test rather than a one-tailed test, the null hypothesis would have been rejected at either level of significance or p-value. This illustrates how changing the directional nature of the hypothesis (one-tailed versus two-tailed test) can result in statistically significant findings.

Researchers can also increase the *sample size*, hence degrees of freedom, and achieve statistically significant research results. If we increase our sample sizes to 100 boys and 100 girls, we enter the t-table with infinity degrees of freedom. The resultant tabled t-values, given a one-tailed test, would be 1.645 at a .05 level of significance or 2.326 at a .01 level of significance. An examination of the t-table further indicates that the tabled t-values are larger for smaller degrees of freedom (smaller sample sizes). The bottom row indicates tabled t-values that are the same as corresponding z-values in the normal distribution given larger sample sizes. This illustrates how increasing the sample size (degrees of freedom greater than 120) can yield a lower tabled t-value for making comparisons to the computed t-value in determining whether the results are statistically significant.

When significance testing, the researcher obtains a sample statistic or "point estimate" of the population parameter. The researcher could compute *confidence intervals* around the sample statistic thereby providing an additional interpretation of the statistical results. The confidence interval width provides valuable information about capturing the population parameter beyond the statistical significance of a "point estimate" of the population value. If the 95% confidence interval for a sample mean ranged from 2.50 to 3.00, then we could conclude with 95% confidence that the interval contained the population mean. Each time we take a random sample of data, the confidence interval would change. If we took all possible samples and computed their confidence intervals, then 95% of the intervals would contain the population mean and 5% would not; therefore, one should not report that the probability is .95 that the interval contains the population mean. Unfortunately, many researchers either do not report confidence intervals and/or misreport them.

Replication of research findings provide support for results obtained. These methods help to address the practical importance of the research study findings. The most meaningful technique would be to *replicate* the study and/or *extend the research* based on earlier findings. This provides the best evidence of research findings or outcomes. Researchers could also use their sample data from a single study and *cross-validate*, *jackknife*, or *bootstrap* the results. In some cases, a researcher might synthesis several findings from research studies by conducting a *meta-analysis*. Most researchers however do not take the time to replicate their study, cross-validate, jackknife, bootstrap, or conduct a meta-analysis. These methods are well known, but not available in most mainstream statistical packages and therefore not readily available to researchers.

Another important consideration above and beyond the significance of a statistical test is the *effect size* or magnitude of difference reported. The interpretation of the effect size can directly indicate whether the statistically significant results are of any practical importance. An example will better illustrate the practical importance of research findings based on an effect size. The previous research study reported a five point average difference between boys and girls in the population on the TAAS test. Is this average five-point difference (approximately getting two test questions correct or incorrect) of practical importance? If we retain the null hypothesis of no difference in the population based on our statistical test of significance, then we conclude that fifth-grade boys and girls achieve about the same. Alternatively, if we reject the null hypothesis in favor of an alternative hypothesis based on our statistical test of significance, then we conclude that fifth-grade boys scored statistically significantly higher on average than the girls at a given level of significance. What are the consequences of our decisions based on a statistical test of significance? If we retain the null hypothesis when it is really false, we make the error of not spending additional money for programs to better educate fifth-grade girls. If we reject the null hypothesis when it is really true, we make the error of spending additional money on programs that are not needed to better educate fifth-grade girls. The effect size helps our practical understanding of the magnitude of the difference detected in a research study. The effect size however should be interpreted based upon a synthesis of findings in several other related studies. This comparison provides a frame of reference for interpreting whether the effect size value is small, medium, or large. The *r* and *d* effect size measures for the computed t-value are computed as follows:

$$r = \sqrt{\frac{t^2}{t^2 + df}}$$

and

$$d = \frac{2t}{\sqrt{df}}$$

The researcher, to achieve significant findings, can manipulate the level of significance, directional nature of the test, and sample size in statistical significance testing. The confidence interval should be reported along with the statistic and p-value to aid in the interpretation of research findings. The effect size helps our practical understanding of the importance of our research results. Replication and/or the extension of a research study are the most meaningful ways to validate findings in a research study. Cross validation, bootstrap, and jackknife methods provide additional information in explaining results from a single study. Results can be statistically significant but have little practical importance.

A few final words of wisdom can be given when faced with significance testing and issues related to the practical importance of research findings. In conducting basic applied research, one asks a question, analyzes data, and answers the research question. Beyond this task, we need to be reminded of several concerns. How do our research findings relate to the research findings in other related research studies? What is the educational importance of our findings? What implications do our research findings have on practice? What recommendations can we make that might affect or modify the underlying theory? What recommendations can we make that might enhance future research efforts?

PRACTICAL R Program

The PRACTICAL program computes an independent *t*-test and outputs the associated values that a researcher should report. The program begins by setting the sample size, first population mean and standard deviation and the second population mean and standard deviation. A random sample of data is then created for two independent samples using the **rnorm** function. The sample means and standard deviations for the two samples are then input into the **t.test** function. The results of the independent *t*-test are output along with the d and r effect size measures.

PRACTICAL Program Output

```
Independent t-test Results

Two sample independent t-test     Sample size=30
Sample One Mean=51.54             Sample One SD=10.42
Sample Two Mean=50.94             Sample Two SD=9.66
t-test=0.23                       df=58
p-value=0.816                     95% Confidence Interval
                                  =-4.588 to 5.798
r Effect=0.031                    d Effect=0.061
```

PRACTICAL Exercises

1. Run the PRACTICAL program 10 times and record the results below.

	t-value	p-value	95%CI	r effect size	d effect size
1)	______	______	______	______	______
2)	______	______	______	______	______
3)	______	______	______	______	______
4)	______	______	______	______	______
5)	______	______	______	______	______
6)	______	______	______	______	______
7)	______	______	______	______	______
8)	______	______	______	______	______
9)	______	______	______	______	______
10)	______	______	______	______	______

a. What conclusions can be drawn about the statistical significance of the computed t-values if the tabled t-value = 1.671 at the .05 level of significance for a one-tailed test?

____

____

b. How many p-values are less than the .05 level of significance? ____________

c. What percent of the confidence intervals captured the population mean difference of zero?________________________

d. What interpretation would you give for the *r* effect size measures for these 10 replications?

____

____

e. What interpretation would you give for the *d* effect size measures for these 10 replications?

____

____

True or False Questions

Meta-Analysis

T F a. Meta-analysis uses subjective techniques to combine research studies.
T F b. The p-value approach combines research findings using chi-square values.
T F c. Various statistics are converted to a common metric so research findings across studies can be quantitatively compared.
T F d. The *r* effect size measure can be interpreted by using a standard reference scale.
T F e. The *d* effect size measure is interpreted relative to findings from a large body of research in an academic discipline.
T F f. When combining effect size measures, it is important to weight by sample size.
T F g. The p-value, log(p), and chi-square approach yield similar results.
T F h. Gene Glass is recognized as creating the term "Meta-Analysis".

Statistical Versus Practical Significance

T F a. Significance testing is the *only* way to know if your findings are important.
T F b. Replication is the *least* meaningful way to determine the validity of your research findings.
T F c. Cross-validation, jackknife, and bootstrap methods provide important information about results when analyzing a single sample of data.
T F d. Research findings can be statistically significant, but have no practical importance to the field of study.
T F e. Increasing the sample size can *always* make a statistical test significant at a given level of significance.
T F f. The sample statistic, p-value, confidence interval, and effect size are recommended values that should be reported in a research study.