Course name:
Empirical software engineering DAT246/DIT278 (formerly DAT245)

Date:
August 27, 2019

Time:
08:30 –12:30

Responsible teachers:
Richard Torkar, David Mattos and Lucas Gren.

Teacher will visit the exam room.
Twice (at 09:30 and at 11:00)

Telephone number:
0723526147 (David Mattos)

Examiner:
Richard Torkar

Aids:
Approved calculator by Chalmers.

Grades:
Maximum points: 49
Chalmers:
Grade 3: [24.5, 31.85)
Grade 4: [31.85, 41.65)
Grade 5: [41.65, 49]
GU:
Grade G: [24.5, 41.65)
Grade VG: [41.65, 49]

Appeal of Grading:
September 17, 2019 at 13.00 – 14:00 (Jupiter, floor 4, room 457, David's office)

Time and place for presentation of solutions:
September 17, 2019 at 13.00 – 14:00 (Jupiter, floor 4, room 457, David's office)

“In science, progress is possible. In fact, if one believes in Bayes' theorem, scientific progress is inevitable as predictions are made and as beliefs are tested and refined.”

— Nate Silver

Text extracted from https://www.nature.com/articles/d41586-019-01307-2
Nature 568, 435 (2019)
doi: *10.1038/d41586-019-01307-2*
24 April, 2019

Rein in the four horsemen of irreproducibility

Dorothy Bishop describes how threats to reproducibility, recognized but unaddressed for decades, might finally be brought under control

"More than four decades into my scientific career, I find myself an outlier among academics of similar age and seniority: I strongly identify with the movement to make the practice of science more robust. It's not that my contemporaries are unconcerned about doing science well; it's just that many of them don't seem to recognize that there are serious problems with current practices. By contrast, I think that, in two decades, we will look back on the past 60 years — particularly in biomedical science — and marvel at how much time and money has been wasted on flawed research.

How can that be? We know how to formulate and test hypotheses in controlled experiments. We can account for unwanted variation with statistical techniques. We appreciate the need to replicate observations.

Yet many researchers persist in working in a way almost guaranteed not to deliver meaningful results. They ride with what I refer to as the four horsemen of the reproducibility apocalypse: publication bias, low statistical power, P-value hacking and HARKing (hypothesizing after results are known). My generation and the one before us have done little to rein these in.

In 1975, psychologist Anthony Greenwald noted that science is prejudiced against null hypotheses; we even refer to sound work supporting such conclusions as 'failed experiments'. This prejudice leads to publication bias: researchers are less likely to write up studies that show no effect, and journal editors are less likely to accept them. Consequently, no one can learn from them, and researchers waste time and resources on repeating experiments, redundantly.

That has begun to change for two reasons. First, clinicians have realized that publication bias harms patients. If there are 20 studies of a drug and only one shows a benefit, but that is the one that is published, we get a distorted view of drug efficacy. Second, the growing use of meta-analyses, which combine results across studies, has started to make clear that the tendency not to publish negative results gives misleading impressions.

Low statistical power followed a similar trajectory. My undergraduate statistics courses had nothing to say on statistical power, and few of us realized we should take it seriously. Simply, if a study has a small sample size, and the effect of an experimental manipulation is small, then odds are you won't detect the effect — even if one is there.

It is wasteful to conduct studies that are underpowered, but researchers have often treated statisticians who point this out as killjoys. In 1977, Jacob Cohen wrote a

definitive book on the subject; ten years later, another statistician wrote, “Small studies continue to be carried out with little more than a blind hope of showing the desired effect” (R. G. Newcombe Br. Med. J. (Clin. Res. Ed.) 295, 656–659; 1987). In fields such as clinical trials and genetics, funders have forced improvements to working practices by insisting that studies be adequately powered. Other disciplines have yet to catch up.

I stumbled on the issue of P-hacking before the term existed. In the 1980s, I reviewed the literature on brain lateralization (how sides of the brain take on different functions) and developmental disorders, and I noticed that, although many studies described links between handedness and dyslexia, the definition of ‘atypical handedness’ changed from study to study — even within the same research group. I published a sarcastic note, including a simulation to show how easy it was to find an effect if you explored the data after collecting results (D. V. M. Bishop J. Clin. Exp. Neuropsychol. 12, 812–816; 1990). I subsequently noticed similar phenomena in other fields: researchers try out many analyses but report only the ones that are ‘statistically significant’.

This practice, now known as P-hacking, was once endemic to most branches of science that rely on P values to test significance of results, yet few people realized how seriously it could distort findings. That started to change in 2011, with an elegant, comic paper in which the authors crafted analyses to prove that listening to the Beatles could make undergraduates younger (J. P. Simmons et al. Psychol. Sci. 22, 1359–1366; 2011). “Undisclosed flexibility,” they wrote, “allows presenting anything as significant.”

The term HARKing was coined in 1998 (N. L. Kerr Pers. Soc. Psychol. Rev. 2, 196–217; 1998). Like P-hacking, it is so widespread that researchers assume it is good practice. They look at the data, pluck out a finding that looks exciting and write a paper to tell a story around this result. Of course, researchers should be free to explore their data for unexpected findings — but P values are meaningless when taken out of context of all the analyses performed to get them.

The problems are older than most junior faculty members, but new forces are reining in these four horsemen. First, the field of meta-science is blossoming, and with it, documentation and awareness of the issues. We can no longer dismiss concerns as purely theoretical. Second, social media enables criticisms to be raised and explored soon after publication. Third, more journals are adopting the ‘registered report’ format, in which editors evaluate the experimental question and study design before results are collected — a strategy that thwarts publication bias, P-hacking and HARKing. Finally, and most importantly, those who fund research have become more concerned, and more strict. They have introduced requirements that data and scripts be made open and methods be described fully.

I anticipate that these forces will soon gain the upper hand, and the four horsemen might finally be slain.”

Contextualizing the text above and with the discussion in class (from both the statistics part and the research methodology part) answer the questions 1, 2 and 3 below.

Question 1) (6pts)

One of the problems pointed out by Dorothy Bishop in the text above is the HARKing (Hypothesizing After the Result). In the context of controlled experiments, discuss:

a) What is an experimental design? **(2pts)**

b) Why HARKing is a problem? **(2pts)**

c) How experimental design can help researchers not to fall for HARKing **(2pts)**?

Question 2) (6 pts)

P-hacking is a general term that consists of consists of 'try out many analyses but report only the ones that are 'statistically significant''. P-hacking together with the publication bias has particularly strong effect in increasing the number of published false positives. P-hacking is also connected to the alpha inflation phenomenon.

a) What is false positive in statistics? **(1 pt)**
b) How do we minimize false positives in a statistical analysis? **(1 pt)**
c) What is alpha inflation? **(1pt)**
d) How do we reduce the problem of alpha inflation? **(1pt)**
e) Explain the relation between alpha inflation and p-hacking **(2 pt)**

Question 3) (5 pts)

"Small studies continue to be carried out with little more than a blind hope of showing the desired effect".

a) What is an underpowered study? **(1 pt)**
b) "Simply, if a study has a small sample size, and the effect of an experimental manipulation is small, then odds are you won't detect the effect — even if one is there". The author describes a specific type of error in controlled experiments. Which error is this? **(1 pt)**
c) What is the relationship between power and statistical significance? **(1 pt)**
d) What are the 2 ways that we can increase power in experiments? **(1 + 1 pt)**

Question 4) This question is based on the concepts presented at the paper:
"K.-J. Stol and B. Fitzgerald. 2018. The ABC of Software Engineering Research. ACM Trans. Softw. Eng. Methodol. 27, 3, Article 11 (September 2018), 51 pages. DOI: https://doi.org/10.1145/3241743"
What is the difference between a "Field study" and a "Laboratory experiment"? Explain the difference in terms of the Goal of the study, the Context, level of obtrusiveness, the level of generalizability and the precision of the measurement of the behavior. Give a short example of each study. **(0.5 pt for each table slot, total 6pts)**

	Field Study	**Laboratory experiment**
Goal of the study:		
Context:		
Level of obtrusiveness		
Level of generalizability		
Precision of the measurement of the behavior		
Example (short)		

The questions 5 and 6 are based on the publication:
Beatriz Bernárdez, Amador Durán, José A. Parejo, Antonio Ruiz–Cortés,
"An experimental replication on the effect of the practice of mindfulness in conceptual modeling performance", Journal of Systems and Software, 136 153-172, 2018

Abstract:
Context: Mindfulness is a meditation technique aimed to increase clearness of mind and awareness. In the 2013–2014 academic year, an experiment was carried out to test whether the practice of mindfulness during 4 weeks improved or not the conceptual modeling performance using UML class diagrams of 32 second–year students of Software Engineering at the University of Seville.
Objective: An internal replication with some changes in the original design was performed in the first semester of the 2014–2015 academic year in order to confirm the insights provided by the original study and increase the confidence in its conclusions. The sample were 53 students with the same profile than in the original study.
Method: Half the students (27 subjects) practiced mindfulness during 6 weeks, while the other half (26 subjects), i.e. the control group, received no treatment during that time. All the students developed two conceptual models using UML class diagrams from a transcript of an interview, one before and another after the 6 weeks of mindfulness sessions, and the results were compared in terms of conceptual modeling effectiveness and efficiency.
Results: REMOVED FOR PURPOSE OF THIS EXAM
Conclusion REMOVED FOR PURPOSE OF THIS EXAM

Further information from the paper:
The authors consider the following independent variables:

- Training workshop (TRWK): this factor represents the training workshop in which the students participated. It has two levels, *mindfulness* and *public speaking*.
- Conceptual modeling exercise (CMEX): this factor has two levels, pre–exercise and post–exercise, which correspond respectively to the conceptual modeling exercises performed by the students before and after participating in the training workshops.

The dependent variables are:

- Effectiveness: the percentage of semantic quality achieved by a subject measured in a ratio scale
- Effectiveness: the percentage of semantic quality achieved by a subject measured in a ratio scale

The parameters of the experiment are:

- The background of the students in conceptual modeling: students with prior knowledge and practice in conceptual modeling would have performed better in the conceptual modeling exercises than the rest of the students. In order to avoid this situation, we tried to have a sample as homogeneous as possible, discarding all the subjects who were either a repeater student or had previous experience in conceptual modeling.
- The ISEIS (the course Introduction to software engineering and information systems) scheduled lessons taught to the students: in order to avoid any difference in the content and methodology of the ISEIS scheduled lessons

taught to the students—which included an introduction to conceptual modeling using UML class diagrams and some related exercises—all the students had the same professor and the same content was taught to all of them at the same pace.

- The complexity of the conceptual modeling exercises and the order in which they were performed by the students: in order to properly compare the results of the conceptual modeling exercises before and after the training workshop sessions, they had to have a similar complexity and a similar level of familiarity of the students with their problem domains.

Experimental design:

- The authors utilized a mixed factorial design, where the CMEX is a within subjects (repeated measured factor) and the TRWK is a between subjects factor. In this type of design we are interested only in the main effect of the between subjects, the main effect within subjects and the interaction within subjects.

Question 5) (Total 9 pts)

The paper presents the within-subjects mixed-model Anova tables for both dependent variables (Note that in this case the main effect of the TRWK and its error are not represented).

a) What are the null hypotheses for each dependent variable? **(1pt for each)**
b) Complete the tables with the mean-square, the F-value and if it is significant or not at alpha=5%. **(2 pts for each table + 0.5 pt for the significance of each table)**

Table 13
Mixed–model ANOVA of effectiveness in the experiment replication.

Source of variation	Type III sum of squares	Degrees of freedom	Mean square	F–ratio	Significance	η_p^2
CMEX	0.286	1				0.312
CMEX*TRWK	0.036	1				0.054
Error(CMEX)	0.629	51				

Table 14
Mixed–model ANOVA of efficiency in the experiment replication.

Source of variation	Type III sum of squares	Degrees of freedom	Mean square	F–ratio	Significance	η_p^2
CMEX	0.275	1				0.382
CMEX*TRWK	0.076	1				0.146
Error(CMEX)	0.446	51				

c) Interpret the results of these table based on the information given earlier **(1pt each)**.

Question 6) (4 pts)

Before conducting the analysis the authors performed some assumptions tests such as the Shapiro-Wilk test and the Levene test

a) What is the Shapiro-Wilk test? What happens if you have a statistically significant result for this test? What happens if you have a non-statistically significant result for this test ?**(1+0.5 + 0.5pt)**
b) The Levene test analyze homoscedasticity. What is homoscedasticity? **(2pt)**

Question 7) Explain and exemplify Random Sampling, Stratified Sampling and Convenience Sampling **(1 for explanation +1 for example for each. Total 6 pts)**

Question 8) (Total 7 pts)

a) What is the difference between a reliable survey and a survey with high validity **(1 pt)**
b) When measuring the reliability of a survey instrument we mainly use four types of approaches. Name (**0.5 pts each)** and exemplify **(0.5 each)** three of them **(Total 3pts)**
c) We have two common ways to evaluate a survey instrument. Which are they (**0.5 pt each)** and give an example for each illustrating the differences **(1 pt each example) (3pts total)**

APPENDIX A

Standard deviation of the sample mean:

$$\frac{\sigma}{\sqrt{n}}$$

Variance of a sample:

$$s^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$$

Pooled variance of a sample:

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

Z-test statistic:

$$\boldsymbol{z = (\overline{X} - \mu)\ /\sigma}$$

T test statistic:

One-sample t-test

$$\boldsymbol{T = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}}$$

df = *n* - 1

Two-sample t-test

$$\boldsymbol{T = \frac{\overline{X}_1 - \overline{X}_2 - diff}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}}$$

$df = n_1 + n_2 - 2$

One-way analysis of variance structuring help

The commonly used normal linear models for a completely randomized experiment are:

$y_{i,j} = \mu_j + \varepsilon_{i,j}$ (the means model)

or

$y_{i,j} = \mu + \tau_j + \varepsilon_{i,j}$ (the effects model)

where

$i = 1, \ldots, I$ is an index over experimental units

$j = 1, \ldots, J$ is an index over treatment groups

I_j is the number of experimental units in the jth treatment group

$I = \sum_j I_j$ is the total number of experimental units

$y_{i,j}$ are observations

μ_j is the mean of the observations for the jth treatment group

μ is the grand mean of the observations

τ_j is the jth treatment effect, a deviation from the grand mean

$\sum \tau_j = 0$

$\mu_j = \mu + \tau_j$

$\varepsilon \sim N(0, \sigma^2)$, $\varepsilon_{i,j}$ are normally distributed zero-mean random errors.

Means model calculations:

	Lists of Group Observations					
	1	2	⋯	j	⋯	J
1	y_{11}	y_{12}				y_{1J}
2	y_{21}	y_{22}				y_{2J}
3						
⋮						
i				y_{ij}		
⋮						
	$y_{I_1 1}$	⋯				
		$y_{I_2 2}$	⋯			

	Group Summary Statistics						Grand Summary Statistics	
# Observed	I_1	I_2	⋯	I_j	⋯	I_J	# Observed	$I = \sum I_j$
Sum				$\sum_i y_{ij}$			Sum	$\sum_j \sum_i y_{ij}$
Sum Sq				$\sum_i (y_{ij})^2$			Sum Sq	$\sum_j \sum_i (y_{ij})^2$
Mean	m_1	⋯		m_j	⋯	m_J	Mean	m
Variance	s_1^2	⋯		s_j^2	⋯	s_J^2	Variance	s^2

Comparing model to summaries: $\mu = m$ and $\mu_j = m_j$. The grand mean and grand variance are computed from the grand sums, not from group means and variances.

One-way ANOVA for means model:

Source of variation	Sums of squares Explanatory SS[4]	Sums of squares Computational SS[5]	Degrees of freedom DF	Mean square MS	F
Treatments	$\sum_{Treatments} I_j(m_j - m)^2$	$\sum_j \frac{(\sum_i y_{ij})^2}{I_j} - \frac{(\sum_j \sum_i y_{ij})^2}{I}$	$J-1$	$\frac{SS_{Treatment}}{DF_{Treatment}}$	$\frac{MS_{Treatment}}{MS_{Error}}$
Error	$\sum_{Treatments} (I_j - 1)s_j^2$	$\sum_j \sum_i y_{ij}^2 - \sum_j \frac{(\sum_i y_{ij})^2}{I_j}$	$I-J$	$\frac{SS_{Error}}{DF_{Error}}$	
Total	$\sum_{Observations} (y_{ij} - m)^2$	$\sum_j \sum_i y_{ij}^2 - \frac{(\sum_j \sum_i y_{ij})^2}{I}$	$I-1$		

MS_{Error} is the estimate of variance corresponding to σ^2 of the model.

The tables in the course book are based on the effects model:

Table 8.5. Analysis of variance table for one-factor experiments

COMPONENT	SUM OF SQUARES	PERCENTAGE VARIATION	DEGREES OF FREEDOM	MEAN SQUARE	F CALCULATION	F TABLE
Y	$SSY = \sum Y_{ij}^2$		N			
$\bar{Y}_{..}$	$SSO = N\mu^2$		1			
$Y - \bar{Y}_{..}$	$SST = SSY - SSO$	100	N-1			
A	$SSA = \sum r_j \alpha_i^2$	$100\left(\frac{SSA}{SST}\right)$	k-1	$MSA = \frac{SSA}{k-1}$	$\frac{MSA}{MSE}$	$F_{[1-\alpha;a-1,(N-k)]}$
e	$SSE = SST - SSA$	$100\left(\frac{SSE}{SST}\right)$	N-k	$MSE = \frac{SSE}{(N-k)}$		

Table 10.4 Analysis of variance table for two factors

Component	Sum of squares	Degrees of freedom	Mean square	F- Computed.	F-Table
Y	$SSY = \sum Y_{ij}^2$	abr			
$\overline{Y}..$	$SSO = ab\eta\mu^2$	1			
$Y - \overline{Y}..$	SST = SSY– SSO	abr-1			
A	$SSA = br\sum \alpha_i^2$	a-1	$MSA = \frac{SSA}{a-1}$	$\frac{MSA}{MSE}$	$F_{[1-\alpha;(a-1),ab(r-1)]}$
B	$SSB = ar\sum \beta_j^2$	b-1	$MSB = \frac{SSB}{b-1}$	$\frac{MSB}{MSE}$	$F_{[1-\alpha;(b-1),ab(r-1)]}$
AB	$SSAB = r\sum \alpha\beta_{ij}^2$	(a-1)(b-1)	$MSAB = \frac{SSAB}{(a-1)(b-1)}$	$\frac{MSAB}{MSE}$	$F_{[1-\alpha;(a-1)(b-1),ab(r-1)]}$
e	$SSE = \sum e_{ijk}^2$	ab(r-1)	$MSE = \frac{SSE}{ab(r-1)}$		

abr = I (in means model) = the total number of experimental units (people) in the study in this case.
r = number of replications (individuals) in each group.

abr-1 = (a-1) + (b-1) + (a-1)(b-1) + ab(r-1)

Table 9.4. Analysis of variance by one factor and one block variable

Component	Sum of Squares	Degrees of Freedom	Mean Square	F-Computed	F-Table
Y	$SSY = \sum y_{ij}^2$	ab			
$\overline{Y}..$	$SS0 = ab\mu^2$	1			
$Y - \overline{Y}..$	SST = SSY – SS0	ab-1			
A	$SSA = b\sum \alpha_i^2$	a-1	$MSA = \frac{SSA}{a-1}$	$\frac{MSA}{MSE}$	$F_{[1-\alpha;(a-1);(a-1)(b-1)]}$
B	$SSB = a\sum \beta_j^2$	b-1	$MSB = \frac{SSB}{b-1}$		
e	$SSE = \sum e_{ij}^2$	(a-1)(b-1)	$MSE = \frac{SSE}{(a-1)(b-1)}$		

Table 9.5. Analysis of variance by one factor and two block variables

$$SS_{total} = \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ijk} - \overline{y}_{..})^2 = \sum_{i=1}^{p}\sum_{j=1}^{p} y_{ijk}^2 - \frac{y_{...}^2}{p^2} \qquad SS_{row} = \sum_{i=1}^{p} p(\overline{y}_{i..} - \overline{y}_{..})^2 = \sum_{i=1}^{p} \frac{R_i^2}{p} - \frac{y_{...}^2}{p^2}$$

$$SS_{trt} = \sum_{j=1}^{p} p(\overline{y}_{.j.} - \overline{y}_{..})^2 = \sum_{j=1}^{p} \frac{T_j^2}{p} - \frac{y_{...}^2}{p^2} \qquad SS_{col} = \sum_{k=1}^{p} p(\overline{y}_{..k} - \overline{y}_{..})^2 = \sum_{k=1}^{p} \frac{C_k^2}{p} - \frac{y_{...}^2}{p^2}$$

Source of Variation	Sum of Squares	d.f.	Mean Square	F Ratio
Treatments	SS_{trt}	$p-1$	MS_{trt}	$\frac{MS_{trt}}{MS_E}$
Rows	SS_{row}	$p-1$	MS_{row}	
Columns	SS_{col}	$p-1$	MS_{col}	
Error	SS_E	$(p-1)(p-2)$	MS_E	
Total	SS_{total}	p^2-1		

ANOVA for nested design (corrected, i.e. the course book is wrong):

Correction to the formulas

Table 11.4. Analysis of variance for the data of example 12.1

Source of variation	Sum of squares	Degrees of freedom	Mean square	F-Computed	F-Table
A	$SSA = br\sum_{i=1}^{a}(\bar{y}_{i..} - \bar{y}_{...})^2$	a–1	$MSA = \frac{SSA}{a-1}$	MSA	$F_{[1-\alpha;(a-1),ab(r-1)}$
B within A	$SSB(A) = r\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.} - \bar{y}_{i..})^2$	a(b–1)	$MSB(A) = \frac{SSB}{a(b-1)}$	MSB(A)	$F_{[1-\alpha;a(b-1),ab(r-1)}$
Error	$SSE = \sum_{k=1}^{r}\sum_{i=1}^{a}\sum_{j=1}^{b}$	ab(r–1)	$MSE = \frac{SSE}{ab(r-1)}$		
Total	$SST = \sum_{k=1}^{r}\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ijk} - \bar{y}_{...})^2$	abr–1			

The correct SSE is:

$$\sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij\cdot})^2$$

And, to get the F value you of course divide by MSE.

ANOVA for mixed design (Blocked Factorial Design)

Component	Sum of squares	Degrees of freedom	Mean square	F- Computed	F-Table (α=0.99)
Between groups	$SS_{between-groups}$	3			
Groups or ABC	$SSABC = 2^k r\, C_{ABC}^2$	1	$MSABC = SSABC$	$\frac{MSABC}{MSEbetween-groups}$	$F_{[1-\alpha;\, 1,\, 2]}$
Error	$SSE_{between-groups}$ = Difference	2	$MSEbetween-groups = \frac{SSEbetween-groups}{2}$		
Within groups					
A	$SSA = 2^k r\, C_A^2$	$2^k r-4$	$MSA = SSA$	$\frac{MSA}{MSEwithin-groups}$	$F_{[1-\alpha;\, 1,\, 2^k r-10]}$
B	$SSB = 2^k r\, C_B^2$	1	$MSB = SSB$	$\frac{MSB}{MSEwithin-groups}$	
C	$SSC = 2^k r\, C_C^2$	1	$MSC = SSC$	$\frac{MSC}{MSEwithin-groups}$	
AB	$SSAB = 2^k r\, C_{AB}^2$	1	$MSAB = SSAB$	$\frac{MSAB}{MSEwithin-groups}$	
AC	$SSAC = 2^k r\, C_{AC}^2$	1	$MSAC = SSAC$	$\frac{MSAC}{MSEwithin-groups}$	
BC	$SSBC = 2^k r\, C_{BC}^2$	1	$MSBC = SSBC$	$\frac{MSBC}{MSEwithin-groups}$	
Error	$SSE_{within-groups}$ = Difference	$2^k r-10$			

$$R_A^2 = \frac{SS_A}{SS_{Total}}$$

$$R_B^2 = \frac{SS_B}{SS_{Total}}$$

$$R_{AB}^2 = \frac{SS_{AB}}{SS_{Total}}$$ Etc.

APPENDIX B (*t* Distribution)

One Sided	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
Two Sided	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%
1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291

APPENDIX C (*F* Distribution)

TABLE C.5
Upper Percentage Points of the *F* Distribution

df for denominator	α	df for numerator 1	2	3	4	5	6	7	8	9	10	11	12
1	.25	5.83	7.50	8.20	8.58	8.82	8.98	9.10	9.19	9.26	9.32	9.36	9.41
	.10	39.9	49.5	53.6	55.8	57.2	58.2	58.9	59.4	59.9	60.2	60.5	60.7
	.05	161	200	216	225	230	234	237	239	241	242	243	244
2	.25	2.57	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.37	3.38	3.39	3.39
	.10	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.40	9.41
	.05	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4
	.01	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4
3	.25	2.02	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44	2.44	2.45	2.45
	.10	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.22
	.05	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74
	.01	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	27.1
4	.25	1.81	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08	2.08	2.08
	.10	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.91	3.90
	.05	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91
	.01	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.4
5	.25	1.69	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89
	.10	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.28	3.27
	.05	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.71	4.68
	.01	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.96	9.89
6	.25	1.62	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77	1.77	1.77	1.77
	.10	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.92	2.90
	.05	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00
	.01	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72
7	.25	1.57	1.70	1.72	1.72	1.71	1.71	1.70	1.70	1.69	1.69	1.69	1.68
	.10	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.68	2.67
	.05	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57
	.01	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47
8	.25	1.54	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.63	1.63	1.63	1.62
	.10	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.52	2.50
	.05	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28
	.01	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67
9	.25	1.51	1.62	1.63	1.63	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.58
	.10	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.40	2.38
	.05	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07
	.01	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	5.11
10	.25	1.49	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.55	1.54
	.10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.30	2.28
	.05	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91
	.01	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71

TABLE C.5 (continued)

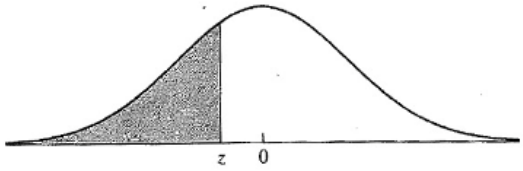
df for denominator	α	*df* for numerator											
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
	.25	1.47	1.58	1.58	1.57	1.56	1.55	1.54	1.53	1.53	1.52	1.52	1.51
11	.10	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.23	2.21
	.05	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79
	.01	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40
	.25	1.46	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.51	1.50	1.50	1.49
12	.10	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.17	2.15
	.05	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69
	.01	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16
	.25	1.45	1.55	1.55	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.47
13	.10	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.12	2.10
	.05	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60
	.01	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96
	.25	1.44	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.46	1.46	1.45
14	.10	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.08	2.05
	.05	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53
	.01	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80
	.25	1.43	1.52	1.52	1.51	1.49	1.48	1.47	1.46	1.46	1.45	1.44	1.44
15	.10	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.04	2.02
	.05	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48
	.01	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67
	.25	1.42	1.51	1.51	1.50	1.48	1.47	1.46	1.45	1.44	1.44	1.44	1.43
16	.10	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	2.01	1.99
	.05	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42
	.01	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.62	3.55
	.25	1.42	1.51	1.50	1.49	1.47	1.46	1.45	1.44	1.43	1.43	1.42	1.41
17	.10	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.98	1.96
	.05	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38
	.01	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.46
	.25	1.41	1.50	1.49	1.48	1.46	1.45	1.44	1.43	1.42	1.42	1.41	1.40
18	.10	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.96	1.93
	.05	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34
	.01	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.43	3.37
	.25	1.41	1.49	1.49	1.47	1.46	1.44	1.43	1.42	1.41	1.41	1.40	1.40
19	.10	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.94	1.91
	.05	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31
	.01	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30
	.25	1.40	1.49	1.48	1.46	1.45	1.44	1.43	1.42	1.41	1.40	1.39	1.39
20	.10	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.92	1.89
	.05	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28
	.01	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23

TABLE C.5 (continued)

df for denominator	α	*df* for numerator											
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
	.25	1.40	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.39	1.38	1.37
22	.10	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.88	1.86
	.05	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23
	.01	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12
	.25	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.38	1.37	1.36
24	.10	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.85	1.83
	.05	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.21	2.18
	.01	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03
	.25	1.38	1.46	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.37	1.36	1.35
26	.10	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.84	1.81
	.05	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15
	.01	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	3.02	2.96
	.25	1.38	1.46	1.45	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34
28	.10	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.81	1.79
	.05	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12
	.01	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.96	2.90
	.25	1.38	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.36	1.35	1.35	1.34
30	.10	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.79	1.77
	.05	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09
	.01	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84
	.25	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.32	1.31
40	.10	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.73	1.71
	.05	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00
	.01	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66
	.25	1.35	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.31	1.30	1.29	1.29
60	.10	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.68	1.66
	.05	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92
	.01	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50
	.25	1.34	1.40	1.39	1.37	1.35	1.33	1.31	1.30	1.29	1.28	1.27	1.26
120	.10	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.62	1.60
	.05	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.87	1.83
	.01	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.40	2.34
	.25	1.33	1.39	1.38	1.36	1.34	1.32	1.31	1.29	1.28	1.27	1.26	1.25
200	.10	2.73	2.33	2.11	1.97	1.88	1.80	1.75	1.70	1.66	1.63	1.60	1.57
	.05	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.84	1.80
	.01	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.34	2.27
	.25	1.32	1.39	1.37	1.35	1.33	1.31	1.29	1.28	1.27	1.25	1.24	1.24
∞	.10	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.57	1.55
	.05	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75
	.01	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.18

APPENDIX D (*Z* Distribution)

TABLE A.2 Cumulative normal distribution (*z* table)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
−3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641