UNIVERSITY OF WATERLOO

MID-TERM EXAMINATION
WINTER TERM 2016

Student Name (Print Legibly) SOLUTIONS
(FAMILY NAME) (GIVEN NAME)

Signature _____

Waterloo Student ID Number _____

COURSE NUMBER	Stat 332
COURSE TITLE	Sampling and Experimental Design
COURSE SECTION	001
DATE OF EXAM	Thursday, February 25, 2016
TIME PERIOD	5:00 PM – 6:20 PM
DURATION OF EXAM	80 minutes
NUMBER OF EXAM PAGES (Including this sheet)	16
INSTRUCTOR	Collin Roberts
EXAM TYPE	Closed Book
ADDITIONAL MATERIALS ALLOWED	Pink Tie calculator

Notes:

1. Do **not** write formulas on this page.
2. Answer all questions in the space provided.
3. If you run out of space, you may write on the blank pages at the end of the exam. Clearly indicate when you have done so.
4. Your grade **will** be influenced by how clearly you express your ideas, and how well you organize your solutions.

Marking Scheme:

Question	Mark	Out of
1		14
2		21
3		11
4		4
Total		50

1. You are testing two brands, 1 and 2, of road paint. You have eight sections of road on which to paint lines. You randomly assign four of the sections to receive each of the two paints. One year after painting, the response variate, namely the visibility of the remaining paint, is measured on a standardized scale. The observed data are recorded here.

Subject	Paint 1	Paint 2
1	50	46
2	51	45
3	46	44
4	47	47

[2] (a) State explicitly whether this experimental design is blocked or un-blocked. Briefly justify your answer.

[6] (b) Let τ_1, τ_2 be the treatment effects of paints 1 and 2, respectively. Construct a 98% confidence interval for the difference $\theta = \tau_1 - \tau_2$. Explain your notation clearly.

Solution: Let Y_{ij} be the response for section j with paint i, with observed value y_{ij}. Assume that $Y_{ij} = \mu + \tau_i + R_{ij}$, where $R_{ij} \sim G(0, \sigma)$ are independent, $i = 1, 2$ and $j = 1, 2, 3, 4$. From the observed data, we obtain

$$\begin{aligned}
\hat{\mu} &= \frac{50+51+46+47+46+45+44+47}{8} = 47 \\
\overline{y}_{1+} &= \frac{50+51+46+47}{4} = 48.5 \\
\overline{y}_{2+} &= \frac{46+45+44+47}{4} = 45.5 \\
\hat{\tau}_1 &= 48.5 - 47 = 1.5 \\
\hat{\tau}_2 &= 45.5 - 47 = -1.5, \text{ so that} \\
\hat{\theta} = \hat{\tau}_1 - \hat{\tau}_2 &= 1.5 - (-1.5) = 3.
\end{aligned}$$

The residual sum of squares is

$$\begin{aligned}
&(50-47-1.5)^2 + (51-47-1.5)^2 + (46-47-1.5)^2 + (47-47-1.5)^2 \\
&+(46-47-(-1.5))^2 + (45-47-(-1.5))^2 + (44-47-(-1.5))^2 + (47-47-(-1.5))^2 \\
= \; & 22.
\end{aligned}$$

Hence we estimate σ by $\hat{\sigma} = \sqrt{\frac{22}{6}} = 1.914854216$, with 6 degrees of freedom. For right-tail probability 0.01 with a t-distribution with 6 degrees of freedom, we select $c = 3.143$. Therefore a 98% confidence interval for the difference in treatment effects is

[6] (c) Determine whether there is evidence that paint 1 is superior to paint 2. Use a threshold of 0.05 to reject your null hypothesis. Explain your notation clearly.

Solution: Keep the notation from part (b). As above, $\hat{\theta} = 3$. Our null hypothesis is $H_0 : \theta = 0$, with alternative hypothesis $H_a : \theta > 0$ (one-sided). Our observed discrepancy measure is

$$\begin{aligned} d &= \frac{\hat{\theta} - 0}{\hat{\sigma}\sqrt{\frac{1}{4} + \frac{1}{4}}} \\ &= \frac{3\sqrt{2}}{1.914854216} \\ &= 2.215646838. \end{aligned}$$

Hence our p-value is $p = P(t_6 \geq 2.215646838)$. From the t-table, we obtain $P(t_6 \geq 1.943) = 0.05$. This implies that $p = P(t_6 \geq 2.215646838) < P(t_6 \geq 1.943) = 0.05$. Therefore we have evidence against the null hypothesis, in other words we have evidence that paint 1 is superior to paint 2.

Marking Instructions:

- Assign one mark each for correctly
  - stating H_0 and H_a
  - applying a one-sided test
  - computing d
  - stating the expression for the p-value
  - explaining why $p < 0.05$
  - interpreting the p-value as evidence against the null hypothesis

2. You are designing an experiment to analyze the effectiveness of food additives at preventing mould in cakes. You have to choose between preservatives 1 and 2, and between red and green food colourings. You bake two cakes for each possible choice, in a randomly assigned order. The response variate is the number of days that pass after the cake is baked and placed in a refrigerator until mould first appears on the cake. The observed data are shown here.

Subject	1R	1G	2R	2G
1	16	11	13	17
2	12	10	11	19

[6] (a) Construct the ANOVA table to test the null hypothesis of no differences among the treatments.

Solution: This model is unblocked and balanced. Use these numbers for the treatments.

Treatment	1	2	3	4
Factor Levels	$1R$	$1G$	$2R$	$2G$

Let Y_{ij} be the number of days for cake j with treatment i, with observed value y_{ij}. Assume that $Y_{ij} = \mu + \tau_i + R_{ij}$, where $R_{ij} \sim G(0, \sigma)$ are independent, $i = 1, 2, 3, 4$ and $j = 1, 2$. From the observed data, we have

$$\begin{aligned}
\hat{\mu} &= \frac{16+12+11+10+13+11}{8} = \frac{73}{8} = 13.625 \\
\overline{y}_{1+} &= \frac{16+12}{2} = \frac{28}{2} = 14 \\
\overline{y}_{2+} &= \frac{11+10}{2} = \frac{21}{2} = 10.5 \\
\overline{y}_{3+} &= \frac{13+11}{2} = \frac{24}{2} = 12 \\
\overline{y}_{4+} &= \frac{17+19}{2} = \frac{36}{2} = 18 \\
\hat{\tau}_1 &= 14 - 13.625 = 0.375, \text{ so } \hat{\tau}_1^2 = 0.140625 \\
\hat{\tau}_2 &= 10.5 - 13.625 = -3.125, \text{ so } \hat{\tau}_2^2 = 9.765625 \\
\hat{\tau}_3 &= 12 - 13.625 = -1.625, \text{ so } \hat{\tau}_3^2 = 2.640625 \\
\hat{\tau}_4 &= 18 - 13.625 = 4.375, \text{ so } \hat{\tau}_4^2 = 19.140625.
\end{aligned}$$

The treatment sum of squares is $0.140625+9.765625+2.640625+19.140625 = 63.375$. The total sum of squares is

$$\begin{aligned}
&(16-13.625)^2 + (12-13.625)^2 + (11-13.625)^2 + (10-13.625)^2 \\
&+(13-13.625)^2 + (11-13.625)^2 + (17-13.625)^2 + (19-13.625)^2 = 75.875.
\end{aligned}$$

Therefore the residual sum of squares is $75.875 - 63.375 = 12.5$, and we obtain the ANOVA table

Source	**Sum of squares**	**Degrees of freedom**	**Mean square (ms)**	**Ratio to residual ms**
Treatments	63.375	3	21.125	6.76
Residual	12.5	4	3.125	
Total	75.875	7		

Marking Instructions:

- Assign one mark for clearly explaining notation introduced.
- Assign one mark each for correct computation (without deducting marks for cascading errors) of
  - treatment sum of squares
  - total sum of squares
  - residual sum of squares (assign zero marks if this is incorrect because of a wrong model choice)
  - treatment mean square and residual mean square (half mark each)
  - ratio of treatment mean square to residual mean square

[3] (b) Determine whether there is evidence of a difference between the treatments. Use a threshold of 0.05 to reject your null hypothesis.

Solution: From the ANOVA table, we have the observed discrepancy measure 6.76. our p-value is $p = P(F_{3,4} \geq 6.76)$. From the F-table, we obtain $P(F_{3,4} \geq 6.59) = 0.05$. This implies

$$p = P(F_{3,4} \geq 6.76) < P(F_{3,4} \geq 6.59) = 0.05.$$

Therefore we have evidence against the null hypothesis of no difference among the treatments. In other words, we have evidence of a difference among the treatments.

Marking Instructions:

- Assign one mark for correctly stating the expression for the p-value.
- Assign one mark for a correct explanation of why $p < 0.05$.
- Assign one mark for the interpretation of the p-value as evidence of a difference.

[6] (c) Determine whether there is evidence of a difference between the treatment effects for preservatives 1 and 2, with the red food colouring. Use a threshold of 0.05 to reject your null hypothesis. Explain your notation clearly.

Solution: We use the contrast

$$\theta = \tau_1 - \tau_3.$$

From the observed data, the estimate is

$$\hat{\theta} = \hat{\tau_1} - \hat{\tau_3} = 0.375 - (-1.625) = 2.$$

The corresponding estimator, $\tilde{\theta}$ satisfies

$$\tilde{\theta} \sim G\left(\theta, \sigma\sqrt{\frac{1}{2} + \frac{1}{2}}\right) = G\left(\theta, \sigma\right),$$

where we estimate σ from the ANOVA table as

$$\hat{\sigma} = \sqrt{3.125} = 1.767766952966369.$$

To test the hypothesis that $\theta = 0$, the observed discrepancy (two-sided) is

$$d = \frac{|\hat{\theta} - 0|}{\hat{\sigma}} = 1.13137085$$

and the p-value is $P(|t_4| \geq 1.13137085)$. From the t-tables, we obtain $P(t_4 \geq 2.776) = 0.025$.

By symmetry this shows that $P(|t_4| \geq 2.776) = 0.05$. Therefore $p = P(|t_4| \geq 1.13137085) > 0.05$. So we have no evidence against the null hypothesis $H_0 : \tau_1 - \tau_3 = 0$.

Marking Instructions:

- Assign one mark for a correct choice of contrast.
- Assign one mark each for correctly applying a two-sided test.
- Assign one mark each for correct computation (without deducting marks for cascading errors) of
  - $\hat{\sigma}$
  - d
- Assign one mark for correctly stating the expression for the p-value.
- Assign one mark for a correct explanation of why $p > 0.05$.

[6] (d) Use an appropriate contrast to test whether there is evidence of an interaction between the factors. Use a threshold of 0.05 to reject your null hypothesis. Explain your notation clearly.

Solution: Use the contrast

$$\theta = (\tau_1 - \tau_2) - (\tau_3 - \tau_4).$$

From the observed data, the estimate is

$$\hat{\theta} = (\hat{\tau_1} - \hat{\tau_2}) - (\hat{\tau_3} - \hat{\tau_4}) = (0.375 - (-3.125)) - ((-1.625) - 4.375) = 9.5.$$

The corresponding estimator, $\tilde{\theta}$ satisfies

$$\tilde{\theta} \sim G\left(\theta, \sigma\sqrt{\frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2}}\right) = G\left(\theta, \sigma\sqrt{2}\right),$$

To test the hypothesis that $\theta = 0$, the observed discrepancy (two-sided) is

$$d = \frac{|\hat{\theta} - 0|}{\hat{\sigma}\sqrt{2}} = 3.8$$

and the p-value is $P(|t_4| \geq 3.8)$. From the t-tables, we obtain $P(t_4 \geq 2.776) = 0.025$. By symmetry this shows that $p = P(|t_4| \geq 2.776) = 0.05$. Therefore $p = P(|t_4| \geq 3.8) < 0.05$. So we have evidence against the null hypothesis of no interaction. In other words, there is evidence of an interaction.

Marking Instructions:

- Assign one mark for a correct choice of contrast.
- Assign one mark each for correctly applying a two-sided test.
- Assign one mark each for correct computation (without deducting marks for cascading errors) of
  - the factor $\sqrt{\frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2}}$ for the standard error
  - d
- Assign one mark for correctly stating the expression for the p-value.
- Assign one mark for a correct explanation of why $p < 0.05$.

3. You are testing five brands of car tires. The response variate is depth of tire tread remaining after one year of driving, measured in millimetres. You have 55 cars to test, each owned and driven by a single driver. You arrange your cars and drivers in groups of five, according to the number of kilometers driven during the previous year. Drivers with large distances driven are grouped together, as are drivers with medium distances driven and drivers with small distances driven.

[2] (a) State explicitly whether this experimental design is blocked or un-blocked. Briefly justify your answer without referring to part (b) of the question.

[6] (b) Using the notation from class, Suppose that the observed data yield the following summary statistics.

- The sample standard deviation of the treatment averages $\overline{y}_{i+}$ (denoted σ_{i+}) equals 0.223828447.
- The sample standard deviation of the block averages $\overline{y}_{+j}$ (denoted σ_{+j}) equals 0.575866146.
- The sample standard deviation of all observed values y_{ij} (denoted σ_{++}) equals 0.738831546.

Construct the ANOVA table to test the hypothesis of no differences among the treatments.

Solution: For a blocked model with t treatments and b blocks, we have that

$$\begin{aligned}
\sigma_{++} &= \sqrt{\frac{\sum_{i,j}(y_{ij}-\overline{y}_{++})^2}{tb-1}} \\
(\sigma_{++})^2 &= \frac{\sum_{i,j}(y_{ij}-\overline{y}_{++})^2}{tb-1} \\
(tb-1)(\sigma_{++})^2 &= \sum_{i,j}(y_{ij}-\overline{y}_{++})^2 \qquad (1)
\end{aligned}$$

$$\begin{aligned}
\sigma_{i+} &= \sqrt{\frac{\sum_{i}(y_{i+}-\overline{y}_{++})^2}{t-1}} \\
(\sigma_{i+})^2 &= \frac{\sum_{i}(y_{i+}-\overline{y}_{++})^2}{t-1} \\
(t-1)(\sigma_{i+})^2 &= \sum_{i}(y_{i+}-\overline{y}_{++})^2 \qquad (2)
\end{aligned}$$

$$\begin{aligned}
\sigma_{+j} &= \sqrt{\frac{\sum_{j}(y_{+j}-\overline{y}_{++})^2}{b-1}} \\
(\sigma_{+j})^2 &= \frac{\sum_{j}(y_{+j}-\overline{y}_{++})^2}{b-1} \\
(b-1)(\sigma_{+j})^2 &= \sum_{j}(y_{+j}-\overline{y}_{++})^2 \qquad (3)
\end{aligned}$$

Now from the given data with $t = 5$ and $b = 11$, we compute

$$\begin{aligned}
\sum_{i,j}(y_{ij} - \overline{y}_{++})^2 &= (tb-1)(\sigma_{++})^2 \\
&= ((5)(11)-1)(0.738831546)^2 \\
&= 29.47709088169651 \qquad (4)\\
b\sum_{i}(y_{i+} - \overline{y}_{++})^2 &= b(t-1)(\sigma_{i+})^2 \\
&= (11)(5-1)(0.223828447)^2 \\
&= 2.204363642203 \qquad (5)\\
t\sum_{j}(y_{+j} - \overline{y}_{++})^2 &= t(b-1)(\sigma_{+j})^2 \\
&= (5)(11-1)(0.575866146)^2 \\
&= 16.58109090544467 \qquad (6)
\end{aligned}$$

and therefore the residual sum of squares is $29.47709088169651 - 2.204363642203 - 16.58109090544467 = 10.69163633404884$.

Therefore the ANOVA table is

Source	**Sum of squares**	**Degrees of freedom**	**Mean square (ms)**	**Ratio to residual ms**
Treatments	2.204363642203	4	0.55109091055075	2.061764516983178
Blocks	16.58109090544467	10		
Residual	10.69163633404884	40	0.267290908351221	
Total	29.47709088169651	54		

[3] (c) Is there evidence of differences among the treatments? Use a threshold of 0.05 to reject your null hypothesis.

Solution: From the ANOVA table, we have the observed discrepancy measure 2.061764516983178. our p-value is $p = P(F_{4,40} \geq 2.061764516983178)$. From the F-table, we obtain $P(F_{4,40} \geq 2.61) = 0.05$. This implies

$$p = P(F_{4,40} \geq 2.061764516983178) > P(F_{4,40} \geq 2.61) = 0.05.$$

Therefore we have **no evidence** against the null hypothesis of no difference among the treatments. In other words, we have no evidence of a difference among the treatments.

[4] 4. Recall the unbalanced model from class. Suppose that we have r_i units for treatment i. We start with the model

$$Y_{ij} = \mu + \tau_i + R_{ij},\ R_{ij} \sim G(0,\sigma), i = 1,\ldots,t,\ j = 1,\ldots,r_i$$

where we now change the constraint to $\sum_i r_i\tau_i = 0$. The ANOVA table for this experimental design can be found on the formula sheet. Prove that the total sum of squares equals the residual sum of squares plus the treatment sum of squares. In other words, prove that the formula

$$\underbrace{\sum_{i,j}(y_{ij} - \overline{y}_{++})^2}_{\text{total}} = \underbrace{\sum_{i,j}(y_{ij} - \overline{y}_{i+})^2}_{\text{residual}} + \underbrace{\sum_i r_i(\overline{y}_{i+} - \overline{y}_{++})^2}_{\text{treatment}}$$

is correct.

Solution:

Now let $1 \le i \le t$ be arbitrary. Then we have

$$\begin{aligned}
\sum_{j=1}^{r_i}(y_{ij} - \overline{y}_{i+}) &= \sum_{j=1}^{r_i} y_{ij} - \overline{y}_{i+}\sum_{j=1}^{r_i} 1 \\
&= \sum_{j=1}^{r_i} y_{ij} - r_i\overline{y}_{i+} \\
&= \sum_{j=1}^{r_i} y_{ij} - \sum_{j=1}^{r_i} y_{ij} \\
&= 0, \text{ completing the proof.}
\end{aligned}$$

You may use the space below for rough work (in which case you may tear off this page), or to continue any other question that you have run out of space answering. In this case, be sure to indicate clearly, in the original location, that the work continues here.

You may use the space below for rough work (in which case you may tear off this page), or to continue any other question that you have run out of space answering. In this case, be sure to indicate clearly, in the original location, that the work continues here.

t-table (right tail)

For each row (degrees of freedom k) and column (right tail probability α), the table entry e satisfies $\Pr(t_k \geq e) = \alpha$. Note that the t-distribution is symmetric about 0.

degrees of freedom	right tail probability 0.25	0.10	0.05	0.025	0.01
1	1.000	3.078	6.314	12.706	31.821
2	0.816	1.886	2.920	4.303	6.965
3	0.765	1.638	2.353	3.182	4.541
4	0.741	1.533	2.132	2.776	3.747
5	0.727	1.476	2.015	2.571	3.365
6	0.718	1.440	1.943	2.447	3.143
7	0.711	1.415	1.895	2.365	2.998
8	0.706	1.397	1.860	2.306	2.896
9	0.703	1.383	1.833	2.262	2.821
10	0.700	1.372	1.812	2.228	2.764
11	0.697	1.363	1.796	2.201	2.718
12	0.695	1.356	1.782	2.179	2.681
13	0.694	1.350	1.771	2.160	2.650
14	0.692	1.345	1.761	2.145	2.624
15	0.691	1.341	1.753	2.131	2.602
16	0.690	1.337	1.746	2.120	2.583
17	0.689	1.333	1.740	2.110	2.567
18	0.688	1.330	1.734	2.101	2.552
19	0.688	1.328	1.729	2.093	2.539
20	0.687	1.325	1.725	2.086	2.528
21	0.686	1.323	1.721	2.080	2.518
22	0.686	1.321	1.717	2.074	2.508
23	0.685	1.319	1.714	2.069	2.500
24	0.685	1.318	1.711	2.064	2.492
25	0.684	1.316	1.708	2.060	2.485
26	0.684	1.315	1.706	2.056	2.479
27	0.684	1.314	1.703	2.052	2.473
28	0.683	1.313	1.701	2.048	2.467
29	0.683	1.311	1.699	2.045	2.462
30	0.683	1.310	1.697	2.042	2.457
35	0.682	1.306	1.690	2.030	2.438
40	0.681	1.303	1.684	2.021	2.423
45	0.680	1.301	1.679	2.014	2.412
50	0.679	1.299	1.676	2.009	2.403
gaussian	0.675	1.282	1.646	1.962	2.330

F-table (right tail) $\alpha = 0.10$

For each row (denominator degrees of freedom *k*) and column (numerator degrees of freedom *j*), the table entry *e* satisfies $P(F(j,k) \geq e) = \alpha$.

denominator degrees of freedom	numerator degrees of freedom											
	1	2	3	4	5	6	7	8	9	10	20	30
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	61.74	62.26
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.44	9.46
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.18	5.17
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.84	3.82
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.21	3.17
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.84	2.80
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.59	2.56
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.42	2.38
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.30	2.25
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.20	2.16
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.12	2.08
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.06	2.01
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.01	1.96
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	1.96	1.91
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	1.92	1.87
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.89	1.84
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.86	1.81
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.84	1.78
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.81	1.76
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.79	1.74
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92	1.78	1.72
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.76	1.70
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89	1.74	1.69
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.73	1.67
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.72	1.66
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.67	1.61
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.61	1.54
50	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.76	1.73	1.57	1.50
100	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.69	1.66	1.49	1.42

F-table (right tail) $\alpha = 0.05$

For each row (denominator degrees of freedom k) and column (numerator degrees of freedom j), the table entry e satisfies $P(F(j,k) \geq e) = \alpha$.

denominator degrees of freedom	numerator degrees of freedom											
	1	2	3	4	5	6	7	8	9	10	20	30
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	248.02	250.10
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.45	19.46
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.66	8.62
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.80	5.75
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.56	4.50
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.87	3.81
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.44	3.38
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.15	3.08
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	2.94	2.86
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.77	2.70
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.65	2.57
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.54	2.47
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.46	2.38
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.39	2.31
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.33	2.25
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.28	2.19
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.23	2.15
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.19	2.11
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.16	2.07
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.12	2.04
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.10	2.01
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.07	1.98
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.05	1.96
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.03	1.94
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.01	1.92
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	1.93	1.84
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.84	1.74
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.78	1.69
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.68	1.57

F-table (right tail) $\alpha = 0.01$

For each row (denominator degrees of freedom *k*) and column (numerator degrees of freedom *j*), the table entry *e* satisfies $P(F(j,k) \geq e) = \alpha$.

denominator degrees of freedom	numerator degrees of freedom											
	1	2	3	4	5	6	7	8	9	10	20	30
1	4052	4999	5404	5624	5764	5859	5928	5981	6022	6056	6209	6260
2	98.50	99.00	99.16	99.25	99.30	99.33	99.36	99.38	99.39	99.40	99.45	99.47
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	26.69	26.50
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.02	13.84
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.55	9.38
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.40	7.23
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.16	5.99
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.36	5.20
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	4.81	4.65
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.41	4.25
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.10	3.94
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	3.86	3.70
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.66	3.51
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.51	3.35
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.37	3.21
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.26	3.10
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.16	3.00
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.08	2.92
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.00	2.84
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	2.94	2.78
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	2.88	2.72
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	2.83	2.67
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	2.78	2.62
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	2.74	2.58
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.70	2.54
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.55	2.39
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.37	2.20
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.27	2.10
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.07	1.89

Formulas:

- **Table 1. ANOVA Table for Completely Randomized Design**
(with t treatments, r observations for each)

Source	**Sum of squares**	**Degrees of freedom**	**Mean square (ms)**	**Ratio to residual ms**
Treatments	$\sum_i r(\overline{y}_{i+} - \overline{y}_{++})^2$	$t-1$	$\frac{\sum_i r(\overline{y}_{i+}-\overline{y}_{++})^2}{t-1}$ (1)	$\frac{(1)}{(2)}$
Residual	$\sum_{i,j}(y_{ij} - \overline{y}_{i+})^2$	$t(r-1)$	$\frac{\sum_{i,j}(y_{ij}-\overline{y}_{i+})^2}{t(r-1)}$ (2)	
Total	$\sum_{i,j}(y_{ij} - \overline{y}_{++})^2$	$tr-1$		

- **Table 3: ANOVA Table for Unbalanced Completely Randomized Design**
(with t treatments, r_i observations for each)

Source	**Sum of squares**	**Degrees of freedom**	**Mean square (ms)**	**Ratio to residual ms**
Treatments	$\sum_i r_i(\overline{y}_{i+} - \overline{y}_{++})^2$	$t-1$	$\frac{\sum_i r_i(\overline{y}_{i+}-\overline{y}_{++})^2}{t-1}$ (1)	$\frac{(1)}{(2)}$
Residual	$\sum_{i,j}(y_{ij} - \overline{y}_{i+})^2$	$\sum_i(r_i-1)$	$\frac{\sum_{i,j}(y_{ij}-\overline{y}_{i+})^2}{\sum_i(r_i-1)}$ (2)	
Total	$\sum_{i,j}(y_{ij} - \overline{y}_{++})^2$	$\sum_i r_i - 1$		

- **Table 5: ANOVA Table for Completely Randomized Block Design**
(with t treatments and b blocks)

Source	**Sum of squares**	**Degrees of freedom**	**Mean square (ms)**	**Ratio to residual ms**
Treatments	$b\sum_i(\overline{y}_{i+} - \overline{y}_{++})^2$	$t-1$	$\frac{b\sum_i(\overline{y}_{i+}-\overline{y}_{++})^2}{t-1}$ (1)	$\frac{(1)}{(2)}$
Blocks	$t\sum_j(\overline{y}_{+j} - \overline{y}_{++})^2$	$b-1$	not interested	
Residual	$\sum_{i,j}(y_{ij} - \overline{y}_{++})^2$ $-b\sum_i(\overline{y}_{i+} - \overline{y}_{++})^2$ $-t\sum_j(\overline{y}_{+j} - \overline{y}_{++})^2$	$(t-1)(b-1)$	$\frac{\text{Residual}}{(t-1)(b-1)}$ (2)	
Total	$\sum_{i,j}(y_{ij} - \overline{y}_{++})^2$	$tb-1$		