Bachelorprüfung SoSe 2023

Fach: Data Science: Ökonometrie

Prüferin: Prof. Regina T. Riphahn, Ph.D.

Vorbemerkungen:

Anzahl der Aufgaben:	Die Klausur besteht aus 3 Aufgaben, die alle bearbeitet werden müssen. **Es wird nur der Lösungsbogen eingesammelt.** Angaben auf dem Aufgabenzettel werden nicht gewertet.
Bewertung:	Es können maximal 60 Punkte erworben werden. Die maximale Punktzahl ist für jede Aufgabe in Klammern angegeben. Sie entspricht der für die Aufgabe empfohlenen Bearbeitungszeit in Minuten.
Erlaubte Hilfsmittel:	• Formelsammlung (ist der Klausur beigefügt) • Tabellen der statistischen Verteilungen (sind der Klausur beigefügt) • Taschenrechner • Fremdwörterbuch
Wichtige Hinweise:	• Sollte es vorkommen, dass die statistischen Tabellen, die dieser Klausur beiliegen, den gesuchten Wert der Freiheitsgrade nicht ausweisen, machen Sie dies kenntlich und verwenden Sie den nächstgelegenen Wert. • Sollte es vorkommen, dass bei einer Berechnung eine erforderliche Information fehlt, machen Sie dies kenntlich und treffen Sie für den fehlenden Wert eine plausible Annahme.

Aufgabe 1: **[21 Punkte]**

Der Datensatz enthält Informationen zu Umsätzen und drei Werbemedien (Youtube, Instagram und Zeitung) von 200 Unternehmen:

$umsatz_i$	= Umsatz von Unternehmen i in Tausend Euro
$youtube_i$	= Werbebudget für Youtube von Unternehmen i in Tausend Euro
$instagram_i$	= Werbebudget für Instagram von Unternehmen i in Tausend Euro
$newspaper_i$	= Werbebudget für Zeitungen von Unternehmen i in Tausend Euro

Hinweis: Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

Sie wollen die Verkäufe auf der Grundlage des ausgegebenen Werbebudgets vorhersagen und entscheiden sich zwischen den folgenden Regressionsmodellen (1-4):

$$umsatz_i = \beta_0 + \beta_1 \cdot youtube_i + u_i \tag{1}$$

$$umsatz_i = \beta_0 + \beta_1 \cdot instagram_i + \beta_2 \cdot instagram_i^2 + u_i \tag{2}$$

$$umsatz_i = \beta_0 + \beta_1 \cdot youtube_i + \beta_2 \cdot instagram_i + \beta_3 \cdot newspaper_i + u_i \tag{3}$$

$$log(umsatz_i) = \beta_0 + \beta_1 \cdot youtube_i + \beta_2 \cdot instagram_i + \beta_3 \cdot newspaper_i + u_i \tag{4}$$

a) Sind Modelle 1 und 2 genestet? Begründen Sie kurz Ihre Antwort. (2 Punkte)

b) Geben Sie den marginalen Effekt von *instagram* im Modell 2 an. Bei welcher Höhe des Werbebudgets für Instagram wird der Umsatz maximiert, wenn $\hat{\beta}_0 = 3,343$, $\hat{\beta}_1 = 0,174$ und $\hat{\beta}_2 = -0,044$? Skizzieren Sie anschließend eine mögliche Beziehung zwischen *instagram* und *umsatz* basierend auf diesen Koeffizientenschätzwerten. Markieren Sie alle bekannten Werte auf den Achsen. (7 Punkte)

c) Sie schätzen das Modell 3 in R und erhalten folgenden Output:

```
              Estimate Std. Error t value  Pr(>|t|)
(Intercept)     3.527      0.374    9.422     0.000
youtube         0.045      0.001   32.809     0.000
instagram       0.189      0.009   21.893     0.000
newspaper      -0.001      0.006   -0.177      0.86

Residual standard error: 2.023 on 196 degrees of freedom
Multiple R-squared:  0.897,       Adjusted R-squared:  0.896
F-statistic: 570.3 on 3 and 196 DF,  p-value: < 0.000
```

Sie möchten nun testen, ob die Variablen *instagram* und *newspaper* gemeinsam signifikant zu der Erklärung vom Umsatz beitragen. Benennen Sie das Testverfahren und formulieren Sie die Null- und Alternativhypothese, berechnen Sie die Teststatistik und bestimmen Sie den kritischen Wert. Kann die Nullhypothese auf dem 5%-Signifikanzniveau abgelehnt werden? (6 Punkte)
Hinweis: Das R^2 aus der Schätzung von Modell 1 ist $0,885$.

d) Sie schätzen nun das Modell 4 und erhalten folgenden Output:

```
                Estimate Std. Error t value Pr(>|t|)
(Intercept)        1.921      0.035  55.574    0.000
youtube            0.003      0.001  23.735    0.000
instagram          0.009      0.001  12.365    0.000
newspaper          0.000      0.001   0.545    0.587

Residual standard error: 0.1868 on 196 degrees of freedom
Multiple R-squared:  0.7998,        Adjusted R-squared:  0.7967
F-statistic: 260.9 on 3 and 196 DF,  p-value: <0.000
```

Interpretieren Sie den Koeffizientenschätzer von β_2 inhaltlich und statistisch. (2 Punkte)

e) Nehmen Sie an, dass der Umsatz in dem Modell 4 nicht in Tausend Euro, sondern in Euro geschätzt wird. Wie verändert sich der Schätzer der Koeffizienten β_0 und β_1? (4 Punkte)

Aufgabe 2: **[15,5 Punkte]**

Sie interessieren sich für die Determinanten von Wohnungseinbrüchen. Es steht Ihnen dafür ein Datensatz mit 2756 Wohnbezirken aus bayerischen Großstädten zur Verfügung. Sie beobachten die folgenden Variablen:

$einbruch_i$	=1, bei Einbruch in den letzten 6 Wochen; =0, sonst
$einw_i$	= Anzahl der EinwohnerInnen im Wohnbezirk i in Hundert
alo_i	= Arbeitslosenquote im Wohnbezirk i (0,00 - 1,00)
$hheink_i$	= Durchschnittliches Haushaltsnettoeinkommen im Wohnbezirk i in Euro

Sie schätzen das folgende lineare Regressionsmodell und erhalten untenstehenden Output:

$$einbruch_i = \beta_0 + \beta_1 \cdot einw_i + \beta_2 \cdot alo_1 + \beta_3 \cdot hheink_i + u_i$$

```
                Estimate Std. Error t value Pr(>|t|)
(Intercept)       35.028      2.045  17.129    0.000
einw               0.016      0.006   4.404    0.000
alo                0.121       ???    3.043    0.002
hheink            -0.014      0.008  -1.620    0.105
---

Residual standard error: 0.8548 on 2721 degrees of freedom
Multiple R-squared:  0.01093, Adjusted R-squared:  0.009839
F-statistic: 10.02 on 3 and 2721 DF,  p-value: 0.0000
```

Runden Sie alle Zahlenangaben auf die dritte Nachkommastelle.

a) Interpretieren Sie $\hat{\beta}_1$ inhaltlich und statistisch. (2 Punkte)

b) Berechnen Sie den Standardfehler von $\hat{\beta}_2$. (1 Punkt)

c) Wie hoch ist die Wahrscheinlichkeit eines Wohnungseinbruches in den letzten sechs Wochen für einen Wohnbezirk mit 1000 Einwohnern, einer Arbeitslosenquote von 0,03 und einem durchschnittlichen Nettohaushaltseinkommen von 2500 Euro? (3 Punkte)

d) Sie vermuten, dass die Anzahl von in den Wohnbezirken lebenden Frauen ($einw\_fem_i$) und Männern ($einw\_male_i$) einen unterschiedlichen Einfluss haben könnten. Sie haben beide Maße vorliegen und nehmen diese zusätzlich in ihr Modell mit auf. Welches Problem tritt bei der Schätzung auf? Wie lässt es sich lösen? Begründen Sie Ihre Antwort knapp. (4 Punkte)

e) Beschreiben Sie allgemein, welche Bedingungen zutreffen müssen, damit ein Problem ausgelassener Variablen vorliegt. Welche Folgen hätte dies für den geschätzten Koeffizienten? Diskutieren Sie, ob es sich bei $\hat{\beta}_1$ um den kausalen Effekt der Einwohnerzahl auf die Einbruchswahrscheinlichkeit handelt. (3,5 Punkte)

f) Nennen Sie zwei Nachteile des linearen Wahrscheinlichkeitsmodells. (2 Punkte)

Aufgabe 3: **[23,5 Punkte]**

Sie interessieren sich weiterhin für die Determinanten von Wohnungseinbrüchen. Es steht Ihnen dafür der gleiche Datensatz wie aus Aufgabe 2 mit 2756 Wohnbezirken zur Verfügung, jedoch verwenden Sie die Häufigkeit der Wohnungseinbrüche im letzten Jahr. Sie beobachten den folgenden Satz von Variablen:

$anzahleinbr_i$ = Anzahl der Einbrüche im Jahr in Wohnbezirk i
$einw_i$ = Anzahl der EinwohnerInnen im Wohnbezirk i in Hundert
alo_i = Arbeitslosenquote im Wohnbezirk i (0,00-1,00)
$hheink_i$ = Durchschnittliches Haushaltsnettoeinkommen im Wohnbezirk i in Euro

Sie schätzen das folgende lineare Regressionsmodell und erhalten untenstehenden Output:

$$anzahleinbr_i = \beta_0 + \beta_1 \cdot einw_i + \beta_2 \cdot alo_i + \beta_3 \cdot hheink_i + u_i$$

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)     0.211     0.010    22.078   0.000
einw_i          0.004     0.002     2.397   0.017
alo_i           0.011     0.004     2.718   0.007
hheink_i        0.009     0.003      ???     ???
---

Residual standard error: 0.4113 on 2721 degrees of freedom
Multiple R-squared:  ???,        Adjusted R-squared:  0.007
F-statistic: 7.039 on 3 and 2721 DF,  p-value: 0.0001
```

Runden Sie alle Zahlenangaben auf die <u>dritte</u> Nachkommastelle.

a) Berechnen Sie das R^2 der Schätzung und interpretieren Sie dieses. *Hinweis:* Das adjustierte Bestimmtheitsmaß beträgt 0,055. Nennen Sie einen Nachteil des R^2. (4 Punkte)

b) Berechnen und interpretieren Sie das 99%-Konfidenzintervall für den geschätzten Koeffizienten der Variable *hheink*. Gehen Sie darauf ein, ob der Koeffizient statistisch signifikant von Null verschieden ist. (5 Punkte)

c) Sie vermuten, dass sich das Modell zwischen Wohnbezirken mit direkter Autobahnanbindung und jenen ohne direkte Autobahnanbindung unterscheidet.

   i. Erläutern Sie kurz das Vorgehen und die Entscheidungslogik des Chow-Tests auf Strukturbruch, den Sie mittels der Variable $autobahn_i$ durchführen können. (3 Punkte)

   ii. Führen Sie einen Chow-Test auf Strukturbruch am 10%-Niveau durch. Geben Sie Hypothesen, Teststatistik, kritischen Wert und Ihre Testentscheidung für das 10%-Signifikanzniveau an. (6 Punkte)
   Hinweise: $SSR_{pooled} = 3159,781$, $SSR_1 = 1603,674$ (für $autobahn = 0$), $SSR_2 = 1550,10$ (für $autobahn = 1$)

d) Welchen Effekt hat Heteroskedastie auf Unverzerrtheit und Effizienz des KQ-Schätzers? (2 Punkte)

e) Benennen Sie den Fehler in folgendem R-Code: (1,5 Punkte)

```
filter(autobahn=1) %>%

ggplot(data = einbruchstat, aes(x=hheink, y=anzahleinbr)) +

geom_point()
```

f) Was wird mit nachfolgender Code-Zeile ausgeführt? (1 Punkt)

```
fit <-  lm(anzahleinbr ~ einw, data=einbruchstat)
```

g) Sie erweitern Ihren Code. Welche Wirkung hat das auf die Ausgabe? (1 Punkt)

```
fit1 <- lm(anzahleinbr ~ einw , data=einbruchstat, subset=female!=1)
```

Formelsammlung – Praxis der empirischen Wirtschaftsforschung

Kapitel 1:

$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = \sum_{i=1}^{n} x_i(y_i-\bar{y})$$

$$= \sum_{i=1}^{n}(x_i-\bar{x})y_i$$

$$= \sum_{i=1}^{n} x_i y_i - n\bar{x}\cdot\bar{y}$$

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$$

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

$Var(aX + bY) = a^2\, Var(X) + b^2\, Var(Y) + 2ab\, Cov(X,Y)$

Für identisch und unabhängig verteilte Zufallsvariablen Y_i:

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i-\bar{Y}_i\right)^2$$

Kapitel 2:

$y_i = \beta_0 + \beta_1 x_i + u_i$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$\sum_{i=1}^{n}\hat{u}_i = 0 \qquad \sum_{i=1}^{n} x_i\hat{u}_i = 0$$

$$SST \equiv \sum_{i=1}^{n}(y_i-\bar{y})^2$$

$$SSE \equiv \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$$

$$SSR \equiv \sum_{i=1}^{n}\hat{u}_i^2 = \sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2$$

$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$, $0 \le R^2 \le 1$

$$E\left(\hat{\beta}_0\right) = \beta_0 \qquad E\left(\hat{\beta}_1\right) = \beta_1$$

$$Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{\sigma^2}{SST_x}$$

$$Var(\hat{\beta}_0) = \frac{\sigma^2 \frac{1}{n}\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$\hat{\sigma}^2 = \frac{1}{(n-2)}\sum_{i=1}^{n}\hat{u}_i^2 = \frac{SSR}{(n-2)}$$

Regression durch den Ursprung:

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

Kapitel 3:

$$R^2 = \frac{SSE}{SST} = \frac{\left(\sum_{i=1}^{n}(y_i-\bar{y})(\hat{y}_i-\bar{\hat{y}})\right)^2}{\left(\sum_{i=1}^{n}(y_i-\bar{y})^2\right)\left(\sum_{i=1}^{n}(\hat{y}_i-\bar{\hat{y}})^2\right)}$$

Wenn $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$

und $\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$

dann $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2\tilde{\delta}_1$ mit $x_2 = \tilde{\delta}_0 + \tilde{\delta}_1 x_1$

Allgemein für j = 1, 2, …, k:

$$Var\left(\hat{\beta}_j\right) = \frac{\sigma^2}{SST_j\left(1-R_j^2\right)}$$

$$SST_j = \sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\hat{u}_i^2}{n-k-1} = \frac{SSR}{n-k-1}$$

$$se(\hat{\beta}_j) = \frac{\hat{\sigma}}{\left[SST_j(1-R_j^2)\right]^{\frac{1}{2}}}$$

MLR.1: Modell der Grundgesamtheit
$y = \beta_0 + \beta_1 x_i + \beta_2 x_2 + \ldots + \beta_k x_k + u$

MLR.2: Zufallsstichprobe der Größe n folgt dem Bevölkerungsmodell.

MLR.3: Keine unabhängige Variable ist konstant. Keine perfekte Kollinearität.

MLR.4: $E\left(u \mid x_1, x_2, \ldots, x_k\right) = 0$

MLR.5: $Var\left(u \mid x_1, x_2, \ldots, x_k\right) = \sigma^2$

MLR.6: u ist von x_1, x_2, …, x_k unabhängig und $u \sim \text{Normal}(0, \sigma^2)$.

Kapitel 4:

$$(\hat{\beta}_j - \beta_j) / se(\hat{\beta}_j) \sim t_{n-k-1}$$

$$-t_{\frac{\alpha}{2},n-k-1} \le \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta}_j)} \le t_{\frac{\alpha}{2},n-k-1}$$

$$\hat{\beta}_j - c \cdot se(\hat{\beta}_j) \le \beta_j \le \hat{\beta}_j + c \cdot se(\hat{\beta}_j)$$

$$F \equiv \frac{(SSR_r - SSR_u)/q}{SSR_u/(n-k-1)}$$

$$F = \frac{(R_u^2 - R_r^2)/q}{(1-R_u^2)/(n-k-1)}$$

Kapitel 5:

$$\lim_{n\to\infty} P(\left|\hat{\beta}_1 - \beta_1\right| > \varepsilon) \to 0$$

$$\text{plim}\left(\hat{\beta}_1\right) = \beta_1$$

Kapitel 6:

Standardisierung:

$$\frac{y_i - \bar{y}}{\hat{\sigma}_y} = \hat{\beta}_1 \left(\frac{\hat{\sigma}_1}{\hat{\sigma}_y}\right)\left(\frac{x_{i1} - \bar{x}_1}{\hat{\sigma}_1}\right) + \hat{\beta}_2 \left(\frac{\hat{\sigma}_2}{\hat{\sigma}_y}\right)\left(\frac{x_{i2} - \bar{x}_2}{\hat{\sigma}_2}\right) + \ldots + \hat{\beta}_k \left(\frac{\hat{\sigma}_k}{\hat{\sigma}_y}\right)\left(\frac{x_{ik} - \bar{x}_k}{\hat{\sigma}_k}\right) + \frac{\hat{u}_i}{\hat{\sigma}_y}$$

Semielastizität:

$$\%\Delta\hat{y} = 100 \cdot [\exp(\hat{\beta}_j \Delta x_j) - 1]$$

$$\bar{R}^2 = 1 - \frac{SSR/(n-k-1)}{SST/(n-1)} = 1 - \frac{\hat{\sigma}^2}{SST/(n-1)}$$

$$\bar{R}^2 = 1 - (1-R^2)\frac{n-1}{n-k-1}$$

$$P[\hat{y}^0 - t_{\frac{\alpha}{2},n-k-1} \cdot se(\hat{e}^0) \le y^0 \le \hat{y}^0 + t_{\frac{\alpha}{2},n-k-1} \cdot se(\hat{e}^0)] = 1\text{-}\alpha$$

$$\widehat{\log y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \ldots + \hat{\beta}_k x_k$$

$$E(y \mid \mathbf{x}) = \exp(\sigma^2/2) \cdot \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)$$

Kapitel 7:

Regression nach Gruppen

- Modell gepoolt: $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u$

Chow-Test (mit $SSR_P = SSR_{\text{gepooltes Modell}}$):

$$F = \frac{(SSR_P - (SSR_1 + SSR_2))/(k+1)}{(SSR_1 + SSR_2)/\left(n - 2(k+1)\right)}$$

TABLE G.2

Critical Values of the t Distribution

		Significance Level				
	1-Tailed:	**.10**	**.05**	**.025**	**.01**	**.005**
	2-Tailed:	**.20**	**.10**	**.05**	**.02**	**.01**
Degrees of Freedom	1	3.078	6.314	12.706	31.821	63.657
	2	1.886	2.920	4.303	6.965	9.925
	3	1.638	2.353	3.182	4.541	5.841
	4	1.533	2.132	2.776	3.747	4.604
	5	1.476	2.015	2.571	3.365	4.032
	6	1.440	1.943	2.447	3.143	3.707
	7	1.415	1.895	2.365	2.998	3.499
	8	1.397	1.860	2.306	2.896	3.355
	9	1.383	1.833	2.262	2.821	3.250
	10	1.372	1.812	2.228	2.764	3.169
	11	1.363	1.796	2.201	2.718	3.106
	12	1.356	1.782	2.179	2.681	3.055
	13	1.350	1.771	2.160	2.650	3.012
	14	1.345	1.761	2.145	2.624	2.977
	15	1.341	1.753	2.131	2.602	2.947
	16	1.337	1.746	2.120	2.583	2.921
	17	1.333	1.740	2.110	2.567	2.898
	18	1.330	1.734	2.101	2.552	2.878
	19	1.328	1.729	2.093	2.539	2.861
	20	1.325	1.725	2.086	2.528	2.845
	21	1.323	1.721	2.080	2.518	2.831
	22	1.321	1.717	2.074	2.508	2.819
	23	1.319	1.714	2.069	2.500	2.807
	24	1.318	1.711	2.064	2.492	2.797
	25	1.316	1.708	2.060	2.485	2.787
	26	1.315	1.706	2.056	2.479	2.779
	27	1.314	1.703	2.052	2.473	2.771
	28	1.313	1.701	2.048	2.467	2.763
	29	1.311	1.699	2.045	2.462	2.756
	30	1.310	1.697	2.042	2.457	2.750
	40	1.303	1.684	2.021	2.423	2.704
	60	1.296	1.671	2.000	2.390	2.660
	90	1.291	1.662	1.987	2.368	2.632
	120	1.289	1.658	1.980	2.358	2.617
	∞	1.282	1.645	1.960	2.326	2.576

Examples: The 1% critical value for a one-tailed test with 25 *df* is 2.485. The 5% critical value for a two-tailed test with large (> 120) *df* is 1.96.

Source: This table was generated using the Stata® function invttail.

TABLE G.3a

10% Critical Values of the F Distribution

	Numerator Degrees of Freedom									
Denominator Degrees of Freedom	1	2	3	4	5	6	7	8	9	10
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71
90	2.76	2.36	2.15	2.01	1.91	1.84	1.78	1.74	1.70	1.67
120	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60

Example: The 10% critical value for numerator $df = 2$ and denominator $df = 40$ is 2.44.

Source: This table was generated using the Stata® function invFtail.

TABLE G.3b

5% Critical Values of the F Distribution

Denominator Degrees of Freedom	Numerator Degrees of Freedom 1	2	3	4	5	6	7	8	9	10
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Example: The 5% critical value for numerator $df = 4$ and large denominator df (∞) is 2.37.

Source: This table was generated using the Stata® function invFtail.