Lehrstuhl für Statistik und empirische Wirtschaftsforschung
Prof. Regina T. Riphahn, Ph.D.

Bachelorprüfung WS 2018/2019

Fach: Praxis der empirischen Wirtschaftsforschung

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

Vorbemerkungen:

Anzahl der Aufgaben:	Die Klausur besteht aus 4 Aufgaben, die alle bearbeitet werden müssen. **Es wird nur der Lösungsbogen eingesammelt.** Angaben auf dem Aufgabenzettel werden nicht gewertet.
Bewertung:	Es können maximal 90 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: **[27 Punkte]**

Sie analysieren die Lieferzeiten von 1000 Gütern, die aus Großbritannien importiert werden. Alle Güter werden entweder von Dover nach Hamburg verschifft oder von London per Flugzeug nach Hamburg geschickt. Folgende Informationen liegen Ihnen vor:

$Dauer_i$	durchschnittliche Lieferzeit von Gut *i* nach Hamburg in Tagen
$Dover\_Entfern_i$	Entfernung Herstellungsort von Gut *i* - Hafen von Dover in km
$London\_Entfern_i$	Entfernung Herstellungsort Gut *i* - Flughafen von London in km
$Schiff_i$	=1 wenn Gut *i* verschifft wird; =0 sonst
$Flugzeug_i$	=1 wenn Gut *i* per Flugzeug verschickt wird; =0 sonst
$Kuehlgut_i$	=1 wenn es sich bei Gut *i* um Kühlgut handelt; =0 sonst
$Beschraenkung_i$	=1 wenn Gut *i* nach dem Brexit Handelsbeschränkungen unterliegt; =0 sonst

Sie schätzen das folgende lineare Regressionsmodell mit SPSS (Modell I):

$$Dauer_i = \beta_0 + \beta_1\, Dover\_Entfern_i + \beta_2\, London\_Entfern_i + \beta_3\, Schiff_i + \beta_4\, Kuehlgut_i$$
$$+\beta_5\, Dover\_Entfern_i \times Schiff_i + \beta_6\, London\_Entfern_i \times Flugzeug_i + u_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(Konstante)	25,007	3,322	7,53	0,000
Dover_Entfern	-0,007	0,023	-0,30	0,766
London_Entfern	0,025	0,014	1,81	0,070
Schiff	7,295	3,715	1,96	0,050
Kuehlgut	-13,355	3,047	-4,38	0,000
Dover_Entfern × Schiff	0,086	0,026	3,33	0,001
London_Entfern × Flugzeug	0,037	3,322	7,53	0,231

a. Abhängige Variable: Dauer

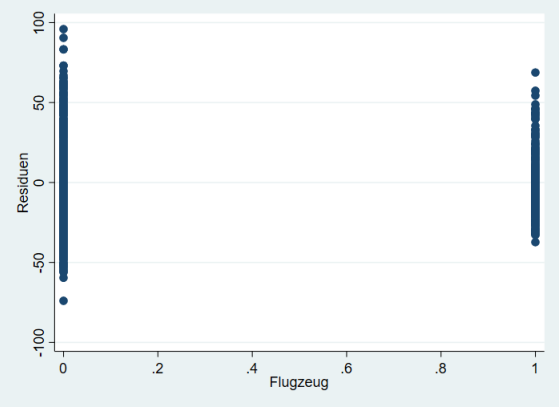
Runden Sie stets auf die dritte Nachkommastelle.

a) Interpretieren Sie $\hat{\beta}_2$ inhaltlich und statistisch. (2,5 Punkte)

b) Berechnen Sie die erwartete Lieferdauer für Kühlgut von einem Herstellungsort, der 70 km von London und 130 km von Dover entfernt ist und das mit dem Flugzeug transportiert wird. (2 Punkte)

c) Sie wundern sich über den positiven und signifikanten Koeffizienten der Variable $London\_Entfern_i$. Nennen Sie eine mögliche ausgelassene Variable, welche die Ursache für eine Überschätzung von β_2 sein könnte, und verdeutlichen Sie am Beispiel die beiden Bedingungen, die dafür erfüllt sein müssen. (4 Punkte)

d) Eine Betrachtung der Residuen kann Aufschluss über die Störterme geben. Betrachten Sie folgende Grafik. Welche der Annahmen des linearen Modells könnte verletzt sein? Begründen Sie Ihre Antwort kurz. (1,5 Punkte)



e) Interpretieren Sie $\hat{\beta}_5$ inhaltlich. (1,5 Punkte)

f) Sie möchten testen, ob die Lieferzeit von Kühlgut mehr als 13 Tage kürzer ist als die Lieferzeit anderer Güter. Führen Sie einen entsprechenden Test durch. Geben Sie Testverfahren, Hypothesen, Teststatistik, Freiheitsgrade, kritischen Wert und Ihre Testentscheidung für das 5%-Signifikanzniveau an. (4,5 Punkte)

g) Berechnen Sie den marginalen Effekt der Variable $Dover\_Entfern_i$ für Güter, die verschifft werden. (2 Punkte)

h) Sie vermuten, dass sich das Modell zwischen Gütern, die von Handlungsbeschränkungen nach dem Brexit betroffen sind, und jenen, die davon nicht betroffen sind, unterscheidet.

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

   ii. Führen Sie einen Chow-Test auf Strukturbruch am 1%-Niveau durch. Geben Sie Hypothesen, Teststatistik, kritischen Wert und Ihre Testentscheidung für das 1%-Signifikanzniveau an. (6 Punkte)
   Hinweise: $SSR_{pooled} = 3232,781$, $SSR_1 = 1296,674$ (für $Beschraenkung = 0$), $SSR_2 = 1498,210$ (für $Beschraenkung = 1$)

Aufgabe 2: **[12,5 Punkte]**

Sie interessieren sich für die Determinanten des finanziellen Erfolgs von Kinofilmen. Hierzu stehen ihnen Daten von *boxofficemojo.com* und *imdb.com* zu Kinostarts von Filmen im Jahr 2018 zur Verfügung, die folgende Informationen beinhalten:

$gross_i$	Umsatz von Film i in Millionen Dollar
$rating_i$	Publikumsbewertung des Films i auf einer Skala von 1-10, wobei 10 die beste zu erreichende Punktzahl darstellt
$theaters_i$	Anzahl der Kinosäle, in denen Film i startete
$opening_i$	Umsatz von Film i am Startwochende in Millionen Dollar

Zunächst beschränken Sie sich auf die 5 erfolgreichsten Filme 2018. Sie unterstellen folgendes Regressionsmodell:

$$gross_i = \beta_0 + \beta_1 rating_i + u_i$$

Film	Umsatz in Millionen Dollar	Rating
Black Panther	700	7,4
Avengers: Infinity War	678	8,5
Incredibles 2	609	7,8
Jurassic World: Fallen Kingdom	418	6,2
Deadpool 2	318	7,8

a) Berechnen Sie $\hat{\beta}_0$ und $\hat{\beta}_1$. Interpretieren Sie $\hat{\beta}_1$ inhaltlich. (6,5 Punkte)

Hinweise: $\sum_{i=1}^{5} x_i y_i = 20765,2$

Falls Sie zu keinem Ergebnis gekommen sind, nehmen Sie $\hat{\beta}_1 = 75$ *an.*
Runden Sie alle Zwischenschritte auf die zweite Nachkommastelle.

Sie schätzen für eine Stichprobe von 100 Kinostarts das folgende lineare Regressionsmodell mit SPSS:

$$gross_i = \beta_0 + \beta_1 rating_i + \beta_2 theaters_i + \beta_3 opening_i + u_i$$

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(Konstante)	76,914	50,360	1,532	0,130
rating	1,160	3,899	???	???
theaters	0,013	0,013	0,995	0,323
opening	2,505	0,269	9,287	0,000

a. Abhängige Variable: gross

Runden Sie stets auf die dritte Nachkommastelle.

b) Interpretieren Sie $\hat{\beta}_3$ inhaltlich. (1 Punkt)

c) Wie würden sich die geschätzten Koeffizienten verändern, wenn *opening* anstatt in Millionen Dollar in Hunderttausend Dollar gemessen wäre? (1,5 Punkte)

d) Berechnen und interpretieren Sie das 99%-Konfidenzintervall von $\hat{\beta}_1$. Ist $\hat{\beta}_1$ auf dem 1%-Niveau statistisch signifikant von 0 verschieden? (3,5 Punkte)

Aufgabe 3: **[10,5 Punkte]**

Sie interessieren sich für das Abstimmungsverhalten von Abgeordenten des amerikanischen Repräsentantenhauses über einen Haushaltsentwurf. Hierzu liegen Ihnen Daten zu 435 Abgeordneten mit folgenden Informationen vor:

$budget_i$	=1 wenn Abgeordnete/r *i* dem Entwurf zustimmt; =0 sonst
$democrat_i$	=1 wenn Abgeordnete/r *i* Mitglied der demokratischen Partei ist; =0 sonst
$votes_i$	Stimmenanteil in Prozent, mit dem Abgeordnete/r *i* die letzte Wahl gewann (kodiert von 0-100)
age_i	Alter von Abgeordnete/r *i* in Jahren

Sie unterstellen folgendes Regressionsmodell und schätzen dieses mit SPSS:

$$budget_i = \beta_0 + \beta_1 democrat_i + \beta_2 votes_i + \beta_3 age_i + \beta_4 age_i^2 + u_i$$

Modellzusammenfassung

Modell	R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers
1	,754(a)	0,569	???	0,151

a. Einflussvariablen: (Konstante), democrat, votes, age, age2

Koeffizienten[a]

Modell	Nicht standardisierte Koeffizienten			
	RegressionskoeffizientB	Standardfehler	T	Signifikanz
(Konstante)	-2,529	0,173	-14.611	0,000
democrat	0,029	0,015	1,992	0,047
votes	-0,002	0,007	-0,286	0,752
age	0,114	0,006	19,150	0,000
age2	-0,001	0,000	-17,299	0,000

a. Abhängige Variable: budget

Runden Sie stets auf die dritte Nachkommastelle.

a) Interpretieren Sie $\hat{\beta}_1$ und $\hat{\beta}_2$ inhaltlich und statistisch. (4 Punkte)

b) Bei welchem Alter wird die Wahrscheinlichkeit, für den Gesetzesentwurf zu stimmen, maximiert? (2 Punkte)

c) Berechnen Sie das angepasste R^2 des Modells. Wie unterscheidet sich die Interpretation des angepassten R^2 von der Interpretation des R^2? (2,5 Punkte)

d) Nennen Sie zwei Schwächen des linearen Wahrscheinlichkeitsmodells. (2 Punkte)

Aufgabe 4 - MC Fragen **[40 Punkte]**

Bitte geben Sie die zutreffende Antwort **auf Ihrem Multiple-Choice-Lösungsblatt** an. Zu jeder Frage gibt es genau eine richtige Antwort. Für jede korrekt beantwortete Frage erhalten Sie einen Punkt. Falsche Antworten führen nicht zu Punktabzug. Bei mehr oder weniger als einer markierten Antwort auf eine Frage gilt diese als nicht beantwortet. **Angaben auf dem Aufgabenblatt werden nicht gewertet.**

1.	Gegeben ist folgendes Modell: $wage_i = \beta_0 + \beta_1 educ_i + \beta_2 age_i + \beta_3 age_i^2 + u_i$. Welche Aussage ist korrekt?
a	Das Modell ist nicht linear in den Parametern.
b	Aufgrund der quadrierten erklärenden Variable liegt perfekte Multikollinearität vor.
c	Ein linearer Zusammenhang zwischen *age* und *wage* kann durch dieses Modell nicht abgebildet werden.
d	Wenn $\beta_3 < 0$, nimmt der marginale Effekt von *age* auf *wage* mit zunehmendem Wert von *age* ab.

2.	Welche Aussage bezüglich der Annahmen des multiplen linearen Regressionsmodells ist zutreffend?
a	MLR.1 - MLR.5 sind notwendig für Unverzerrtheit der Parameterschätzer.
b	Wenn die Annahmen MLR.1 - MLR.5 zutreffen, ist der Erklärungsgehalt des Modells besonders hoch.
c	Die Annahmen MLR.1-MLR.5 werden auch Gauß-Markow-Annahmen genannt.
d	Alle Annahmen müssen zutreffen, damit eine KQ-Schätzung durchgeführt werden kann.

3.	Wofür steht die Abkürzung BLUE?
a	Binary linear unbiased estimator.
b	Best linear unbiased estimator.
c	Best logarithmic unbiased estimator.
d	Best linear uncorrelated estimator.

4.	Welche Folge hat die Aufnahme einer irrelevanten Variable, die mit keiner anderen erklärenden Variable korreliert ist, in das lineare Regressionsmodell?
a	Perfekte Multikollinearität.
b	Geringere Varianz der Parameterschätzer.
c	Einen geringeren Wert des R^2.
d	Einen geringeren Wert des angepassten R^2.

5.	Sie prüfen im Modell $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$ mittels eines F-Tests, ob x_1 und x_2 gemeinsam signifikant sind. Sie finden heraus, dass $R^2_{unrestringiert} = R^2_{restringiert}$. Welche Aussage ist falsch?
a	x_1 und x_2 haben keinen Erklärungsgehalt für y.
b	Die Nullhypothese kann am 5 %-Niveau nicht verworfen werden.
c	Die Aufnahme von x_1 und x_2 verändert die Beobachtungszahl n nicht.
d	Mit einer steigenden Beobachtungszahl erhöht sich der Wert der F-Statistik.

6.	Gegeben ist folgende Lohnregression: $wage_i = \beta_0 + \beta_1 exper_i + \beta_2 female_i + \beta_3 female_i \cdot exper_i + u_i$. Welcher Term beschreibt den Zusammenhang zwischen den Variablen *exper* und *wage* für Männer?
a	β_1
b	$\beta_3 - \beta_2$
c	$\beta_3 - \beta_1$
d	$\beta_1 - \beta_2$

7.	Was versteht man unter der Dummy Variable Trap?
a	Das Auslassen einer relevanten Dummy-Variable.
b	Die Aufnahme von Dummyvariablen, die alle existierenden Zustände abbilden und das daraus resultierende Problem perfekter Kollinearität bei einer Schätzung mit einer Regressionskonstanten.
c	Die Interaktion von zwei oder mehreren Dummyvariablen.
d	Die Schätzung eines Modells, welches nur Dummyvariablen als unabhängige Variablen enthält.

8.	Wann ist β_1 überschätzt, wenn anstatt des Modells $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + u_i$ das Modell $y_i = \beta_0 + \beta_1 x_{i1} + u_i$ geschätzt wird?
a	Wenn $\beta_2 < 0$ und $Cov(x_1, x_2) > 0$.
b	Wenn $\beta_2 = 0$ und $Cov(x_1, x_2) > 0$.
c	Wenn $\beta_2 > 0$ und $Cov(x_1, x_2) = 0$.
d	Wenn $\beta_2 < 0$ und $Cov(x_1, x_2) < 0$.

9.	Das Bestimmtheitsmaß R^2
a	gibt, anders als das angepasste Bestimmtheitsmaß, nicht den durch das Modell erklärten Anteil der Variation in der abhängigen Variable an.
b	steigt nicht, wenn einem Modell eine Variable mit einem geringen Erklärungsgehalt hinzugefügt wird.
c	beträgt eins, wenn alle Beobachtungen auf der Schätzgeraden liegen.
d	wird nur auf Basis der Residuenquadratsumme berechnet.

10.	Welche Aussage bezüglich Homo- und Heteroskedastie ist korrekt?
a	Der KQ-Schätzer ist auch im Falle von Heteroskedastie effizient.
b	Homoskedastie bedeutet, dass die Varianz des Störterms für alle Beobachtungen konstant ist.
c	Heteroskedastie verzerrt die Parameterschätzer.
d	Homoskedastie bedeutet, dass $Var(u) = 0$.

11.	In einer Regression mit n=150 und k=1 beträgt die erklärte Quadratsumme 1164,82 und die totale Quadratsumme 4283,76. Der Standardfehler der Regression beträgt (gerundet auf die zweite Nachkommastelle)
a	4,59.
b	4,56.
c	2,81.
d	5,38.

12.	Die Varianz des Schätzers $\hat{\beta}_1$ im Modell $y_i = \beta_0 + \beta_1 x_{i1} + u_i$
a	sinkt mit steigendem Standardfehler der Regression.
b	sinkt mit steigender Variation in x_1.
c	fällt mit sinkendem R^2.
d	steigt mit steigender Regressionskonstante.

13.	Im multiplen Regressionsmodell mit Konstante testen Sie alle drei Steigungsparameter auf gemeinsame Signifikanz und erhalten einen Wert der F-Statistik von 2,75. Welche Aussage ist korrekt?
a	Die Nullhypothese kann bei einer Stichprobengröße von 60 am 10%-Niveau verworfen werden.
b	Die Nullhypothese kann bei einer Stichprobengröße von 60 am 5%-Niveau verworfen werden.
c	Die Nullhypothese kann bei einer Stichprobengröße von 120 am 5%-Niveau nicht verworfen werden.
d	Die Nullhypothese kann bei einer Stichprobengröße von 120 am 10%-Niveau nicht verworfen werden.

14.	Folgender Auszug aus einem Datensatz liegt Ihnen vor. Um welche Datenstruktur handelt es sich?

Personennr.	Jahr	Einkommen
1	2015	2200
1	2016	2400
2	2015	2300
2	2016	2100

a	Zeitreihendaten.
b	Querschnittsdaten.
c	Gepoolte Querschnittsdaten.
d	Paneldaten.

15.	Der Zusatz *ceteris paribus* kennzeichnet die Annahme, dass
a	die abhängige Variable konstant gehalten wird.
b	alle relevanten Einflussfaktoren konstant gehalten werden.
c	alle berücksichtigten erklärenden Variablen, nicht aber unberücksichtigte Einflussfaktoren, konstant gehalten werden.
d	alle Standardfehler konstant gehalten werden.

16.	Gegeben ist folgendes Modell: $Arbeitslosenquote_i = \beta_0 + \beta_1 Konsumausgaben_i + \beta_2 log(Investitionen_i) + u_i$. Die Arbeitslosenquote ist in Prozent gemessen. Welche Interpretation ist korrekt?
a	Erhöhen sich die Investitionen um einen Prozentpunkt, so erhöht sich die Arbeitslosenquote c.p.i.M. um $\frac{\beta_2}{100}$ Prozentpunkte.
b	Erhöhen sich die Konsumausgaben um eine Einheit, so erhöht sich die Arbeitslosenquote c.p.i.M. um β_1 Prozent.
c	Erhöhen sich die Investitionen um ein Prozent, so erhöht sich die Arbeitslosenquote c.p.i.M. um $\frac{\beta_2}{100}$ Prozentpunkte.
d	Erhöhen sich die logarithmierten Investitionen um eine Einheit, so erhöht sich die Arbeitslosenquote c.p.i.M. um β_2 Prozent.

17.	Die Schätzung des Modells $y_i = \beta_0 + \beta_1 x_{i1} + u_i$ ergibt $\hat{\beta}_0 = 1,02$. Die t-Statistik eines Signifikanztestes von $\hat{\beta}_0$ beträgt 2,15. Was ist der Standardfehler von $\hat{\beta}_0$ (gerundet auf die zweite Nachkommastelle)?
a	2,19.
b	2,11.
c	1,13.
d	0,47.

18.	Nehmen Sie an, Sie erhalten bei der Schätzung eines linearen multiplen Regressionsmodells ein R^2 von 1. Der Standardfehler der Regression
a	beträgt 1.
b	beträgt 0.
c	kann im vorliegenden Fall nicht berechnet werden.
d	wird im vorliegenden Fall auf Basis des angepassten $\overline{R}^2$ berechnet.

19.	Was erhöht die Effizienz einer KQ-Schätzung?
a	Geringere Variation in den erklärenden Variablen.
b	Höhere Varianz der Störterme.
c	Zusätzliche Beobachtungen.
d	Eine Umskalierung der abhängigen Variable.

20.	Die Annahme normalverteilter Störterme
a	ist für die exakte Gültigkeit von t-Tests erforderlich.
b	wird nicht in Modellen mit logarithmierten erklärenden Variablen getroffen.
c	schließt Homoskedastie aus.
d	erhöht die erklärte Quadratsumme.

21.	Welche Aussage bezüglich Interaktionstermen ist richtig?
a	Interaktionsterme können genutzt werden, um Gruppenunterschiede zu untersuchen.
b	Im linearen Wahrscheinlichkeitsmodell ist die Schätzung von Koeffizienten für Interaktionsterme nicht möglich.
c	Interaktionsterme können nur für kontinuierliche Variablen berechnet werden.
d	Interaktionsterme senken das R^2 der Schätzung.

22.	Sie schätzen das Modell $log(arbeitszeit_i) = \beta_0 + \beta_1 bildung_i + \beta_2 berufserfahrung_i + u_i$. Sie vermuten, dass ein Problem ausgelassener Variablen vorliegt, da die Variable Mann nicht berücksichtigt wurde. Welche Aussage ist korrekt?
a	Das Auslassen von Mann in der Schätzung führt zu Homoskedastie.
b	Wenn $Cov(Mann_i, bildung_i) = 0$ und $Cov(Mann_i, arbeitszeit_i) > 0$, dann ist β_1 verzerrt geschätzt.
c	Wenn $Cov(Mann_i, berufserfahrung_i) > 0$ und $Cov(Mann_i, arbeitszeit_i) = 0$, dann ist β_2 unterschätzt.
d	Wenn $Cov(Mann_i, bildung_i) > 0$ und $Cov(Mann_i, arbeitszeit_i) < 0$, dann ist β_1 unterschätzt.

23.	Sie interessieren sich für die Determinanten von Stundenlöhnen und stellen folgendes Modell auf: $lohn_i = \beta_0 + \beta_1 frau_i + \beta_2 alter_i + \beta_3 alter_i^2 + \varepsilon_i$. Welche Aussage ist korrekt?
a	β_1 beschreibt den Lohnunterschied zwischen Männern und Frauen als Funktion des Alters.
b	β_3 ist der marginale Effekt des Alters auf den Lohn.
c	Der mittlere Lohn für Frauen im Alter von 35 Jahren ist $\beta_0 + \beta_2 \cdot 35$.
d	Keine der Antwortmöglichkeiten a-c ist korrekt.

24.	Gegeben sei die Regression $y_i = \beta_0 + \beta_1 x_i + u_i$. Sie schätzen dieses Modell mit einer Stichprobe mit 1722 Beobachtungen und erhalten $\hat{\beta}_1 = 1,02$ und $se(\hat{\beta}_1) = 0,30$. Welche Aussage ist korrekt?
a	Das angepasste R^2 der Regression würde durch das Auslassen von x_i steigen.
b	Es liegen zu wenige Beobachtungen vor, um kausale Aussagen treffen zu können.
c	$\hat{\beta}_1$ ist auf dem 10%-Niveau nicht statistisch signifikant von Null verschieden.
d	$\hat{\beta}_1$ ist auf dem 5%-Niveau statistisch signifikant von Null verschieden.

25.	Welche der folgenden Aussagen bezüglich Dummy-Variablen ist korrekt?
a	Die Koeffizienten von Dummy-Variablen haben keine sinnvolle Interpretation.
b	Dummy-Variablen können verwendet werden, um Mittelwertunterschiede verschiedener Gruppen zu analysieren.
c	Dummy-Variablen können nicht miteinander interagiert werden.
d	Dummy-Variablen können nicht als abhängige Variablen in KQ-Regressionen verwendet werden.

26.	Sie schätzen das Modell $Lohn_i = \beta_0 + \beta_1 bildung_i + u_i$, wobei *Lohn* den Stundenlohn in Euro misst. Das R^2 der Regression beträgt 0,112. Wird der Stundenlohn in Cent anstatt Euro gemessen, dann
a	bleibt das R^2 der Schätzung unverändert.
b	kann das Modell aufgrund perfekter Multikollinearität nicht geschätzt werden.
c	erhöht sich die Wahrscheinlichkeit, dass Heteroskedastie vorliegt.
d	bewirkt dies nur eine Veränderung des angepassten R^2.

27.	Sie schätzen das Modell $stundenlohn_i = \beta_0 + \beta_1 frau_i + \beta_2 abitur_i + u_i$ (wobei der Stundenlohn in Euro gemessen ist) und erhalten $\hat{\beta}_0 = 11$, $\hat{\beta}_1 = -2$ und $\hat{\beta}_2 = 2$. Welche Aussage ist im Rahmen des Modells korrekt?
a	Männer mit Abitur verdienen c.p.i.M. 11 Euro.
b	Ein Abitur lohnt sich für Frauen mehr als für Männer.
c	Eine Frau mit Abitur verdient genauso viel wie ein Mann ohne Abitur.
d	Die Konstante gibt den durchschnittlichen Stundenlohn für Frauen ohne Abitur an.

28.	Im linearen Wahrscheinlichkeitsmodell
a	können die vorhergesagten Werte für *y* größer als 1 sein.
b	liegt zwangsläufig Homoskedastie vor.
c	können keine logarithmierten Variablen als Regressoren verwendet werden.
d	liegen die vorhergesagten Werte für *y* im Intervall [-1,1].

29.	Welche Aussage bezüglich der Interpretation eines log-log-Modells ist richtig?
a	Durch die Logarithmierung von Variablen erhalten Ausreißer ein größeres Gewicht in der Schätzung.
b	Die geschätzten Koeffizienten können als Elastizitäten interpretiert werden.
c	Die Koeffizienten eines solchen Modells können nicht sinnvoll interpretiert werden.
d	Keine der Antwortmöglichkeiten a-c ist korrekt.

30.	Ein Koeffizient ist auf dem 5%-Niveau statistisch signifikant. Welche Aussage ist richtig?
a	Der Koeffizient ist auch auf dem 10%-Niveau statistisch signifikant.
b	Der Koeffizient bildet einen kausalen Effekt ab.
c	Der Koeffizient ist auch auf dem 1%-Niveau statistisch signifikant.
d	Da der Koeffizient statistisch signifikant ist, liegt kein Problem mit Heteroskedastie vor.

31.	Sie schätzen das Modell $Gehalt_i = \beta_0 + \beta_1 Vollzeit_i + \beta_2 Teilzeit_i + u_i$ für 944 Voll- und Teilzeitangestellte, wobei $Teilzeit_i$ und $Vollzeit_i$ Dummy-Variablen sind, die anzeigen, ob Person i in Teilzeit, bzw. Vollzeit arbeitet. Welche Aussage ist richtig?
a	Aufgrund der großen Fallzahl ist Multikollinearität kein Problem.
b	Im vorliegenden Fall wird ein lineares Wahrscheinlichkeitsmodell geschätzt.
c	Es liegt Homoskedastie vor.
d	Aufgrund perfekter Multikollinearität kann das Modell in der vorliegenden Form nicht geschätzt werden.

32.	Sie schätzen das Modell $Lohn_i = \beta_0 + \beta_1 migrant_i + \beta_2 log(arbeitszeit_i) + u_i$ und erhalten $\hat{\beta}_1 = -2,307$ und $\hat{\beta}_2 = 4,100$, wobei der Lohn in Euro gemessen ist. Welche der Aussagen ist korrekt?
a	Eine Erhöhung der Arbeitszeit um 1% führt c.p.i.M. zu einem um 0,041 Euro höheren Lohn.
b	Eine Erhöhung der Arbeitszeit um 1 Stunde führt c.p.i.M. zu einem um 4,100 Euro höheren Lohn.
c	β_2 kann nicht sinnvoll interpretiert werden.
d	Eine Erhöhung der Arbeitszeit um 1% führt c.p.i.M. zu einem um 4,100 Euro höheren Lohn.

33.	Mithilfe eines F-Tests kann untersucht werden,
a	ob erklärende Variablen umskaliert werden sollten.
b	ob mehrere Regressoren gemeinsam signifikant sind.
c	ob die abhängige Variable dichotomisiert werden sollte.
d	ob die abhängige Variable logarithmiert werden sollte.

34.	Das 99%-Konfidenzintervall eines Koeffizienten $\hat{\beta}_1$ lautet [0,020;0,060]. Welche der Aussagen ist korrekt?
a	Es können keine Aussagen über $\hat{\beta}_1$ getroffen werden.
b	$\hat{\beta}_1$ ist auf dem 5%-Niveau statistisch signifikant von Null verschieden.
c	Es kann keine Aussage darüber getroffen werden, ob $\hat{\beta}_1$ auf dem 10%-Niveau statistisch signifikant von Null verschieden ist.
d	$\hat{\beta}_1 = 0,030$.

35.	Was senkt die Effizienz einer Schätzung?
a	Das Vorliegen von Homoskedastie.
b	Die Aufnahme relevanter Variablen mit hohem Erklärungsgehalt.
c	Die Aufnahme irrelevanter Regressoren.
d	Die Vergrößerung der Stichprobe.

36.	Welche Aussage trifft für die Schätzung des Modells $Zufriedenheit_i = \beta_0 + \beta_1 westdeutschland_i + \beta_2 frau_i + \beta_3 westdeutschland_i \cdot frau_i + u_i$ zu? Bei $westdeutschland_i$ und $frau_i$ handelt es sich um Dummy-Variablen.
a	β_0 ist die vorhergesagte Zufriedenheit für westdeutsche Männer.
b	$\beta_2 + \beta_3$ misst den Zufriedenheitsunterschied zwischen westdeutschen Männern und westdeutschen Frauen.
c	β_2 ist der mittlere Unterschied in der Lebenszufriedenheit zwischen Männern und Frauen.
d	β_1 ist der mittlere Unterschied in der Lebenszufriedenheit zwischen Ost- und Westdeutschen.

37.	Heteroskedastie
a	beeinflusst die Konfidenzintervalle von Koeffizientenschätzern nicht.
b	hat Einfluss auf Koeffizientenschätzer.
c	hat keinen Einfluss auf die Ergebnisse von t-Tests.
d	Alle Aussagen a-c sind falsch.

38.	Sie schätzen das Modell $Einkommen_i = \beta_0 + \beta_1 verheiratet_i + u_i$, wobei $Einkommen$ das Jahreseinkommen in Euro misst und erhalten $\hat{\beta}_1 = 12451$. Wäre das Einkommen in Tausend Euro anstatt in Euro gemessen, dann
a	lässt sich auf Basis der gegebenen Informationen keine Aussage bezüglich $\hat{\beta}_1$ treffen.
b	wäre $\hat{\beta}_1 = 12,451$.
c	wäre $\hat{\beta}_1 = 0,125$.
d	wäre die Konstante unverändert.

39.	Ein omitted variable bias im Modell $y_i = \beta_0 + \beta_1 x_1 + u_i$
a	liegt immer dann vor, wenn eine relevante erklärende Variable ausgelassen wurde.
b	ist immer positiv, wenn der Zusammenhang zwischen einer ausgelassenen Variable x_2 und der abhängigen Variable positiv ist.
c	liegt vor, wenn eine Variable ausgelassen wurde, die mit x_1 korreliert und keinen Einfluss auf y hat.
d	bedeutet eine Über- oder Unterschätzung von β_1.

40.	Sie schätzen das Modell $log(Lohn_i) = \beta_0 + \beta_1 frau_i + \beta_2 bildung_i + u_i$ und erhalten $\hat{\beta}_1 = -0,107$ und $\hat{\beta}_2 = 0,042$. Welche der Aussagen ist korrekt?
a	In diesem Modell kann β_2 nicht als marginaler Effekt interpretiert werden.
b	Frauen verdienen c.p.i.M. etwa 10,7% weniger als Männer.
c	Frauen verdienen c.p.i.M. etwa 0,107 Euro weniger als Männer.
d	Der Lohnunterschied zwischen Männern und Frauen variiert mit der Bildung.

Formelsammlung – Praxis der empirischen Wirtschaftsforschung

Kapitel 1:

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

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

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

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

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

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

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:

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

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 $x_2=\tilde{\delta}_0+\tilde{\delta}_1x_1$

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: $E\left(u\,|\,x_1,x_2,\ldots,x_k\right)=0$

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

MLR.6: u ist von x_1, x_2, …, x_k unabhängig und u ~ Normal (0, σ^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:

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

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

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:

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

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

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

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

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

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

Kapitel 7:

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

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

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

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

Denominator Degrees of Freedom	Numerator 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.

TABLE G.3c

1% 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	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
90	6.93	4.85	4.01	3.54	3.23	3.01	2.84	2.72	2.61	2.52
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32

Example: The 1% critical value for numerator $df = 3$ and denominator $df = 60$ is 4.13.

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