SCM 651 Fall 2018, Solutions to Individual Assignment
Due by midnight, Tuesday, 10/9/2018; Total Points = 45

1.(6 pt) Suppose you selected a random sample of 50 students at University X and asked them to rate how interested they are in playing video games in their leisure time on a 1-5 scale (1: not interested at all, 5: very interested). You also noted the gender of the respondent. The results are summarized as the following cross-tabulation:

	Interest				
Gender	1	2	3	4	5
Female	6	3	6	3	2
Male	4	2	9	5	10

Suppose you want to test, at a 99% level of confidence, the null hypothesis that there is no association between gender and interest in playing video games. In the present case, is it valid to use the chi-square test with the original cross-tabulation? Clearly state yes or no and why you say so. If no, modify the original cross-tabulation so that a chi-square test becomes valid. Using the original or modified cross-tab as appropriate, perform a chi-square test of the null hypothesis at a 99% level of confidence. (Show work)

Answer: Augmenting the table by row totals and column totals:

	Interest					
Gender	1	2	3	4	5	Row Totals
Female	6	3	6	3	2	20
Male	4	2	9	5	10	30
Column Totals	10	8	15	8	12	$n = 50$

From this table:

$$E_{11} = \frac{20 \times 10}{50} = 4 \quad E_{12} = \frac{20 \times 5}{50} = 2 \quad E_{13} = \frac{20 \times 15}{50} = 6$$

$$E_{14} = \frac{20 \times 8}{50} = 3.2 \quad E_{15} = \frac{20 \times 12}{50} = 4.8$$

$$E_{21} = \frac{30 \times 10}{50} = 6 \quad E_{22} = \frac{30 \times 5}{50} = 3 \quad E_{23} = \frac{30 \times 15}{50} = 9$$

$$E_{24} = \frac{30 \times 8}{50} = 4.8 \quad E_{25} = \frac{30 \times 12}{50} = 7.2$$

$E_{ij} > 1$ in all cells. However, $E_{ij} \geq 5$ in 4 out of 10 cells, that is, in less than 80% of the cells. Hence chi-square test is not valid with the original table and a modification is necessary here.

Example of modification: We combine columns 1 and 2 into new column 1, keep column as new column 2, and combine columns 4 and 5 into new column 3. This gives us:

$$O_{11} = 6 + 3 = 9,\ E_{11} = 4 + 2 = 6 \quad O_{12} = 6,\ E_{22} = 6 \quad O_{13} = 3 + 2 = 5,\ E_{13} = 3.2 + 4.8 = 8$$

$$O_{21} = 4 + 2 = 6,\ E_{21} = 6 + 3 = 9 \quad O_{22} = 9,\ E_{22} = 9 \quad O_{23} = 5 + 10 = 15,\ E_{23} = 4.8 + 7.2 = 12$$

$$\chi^2 = \frac{(9-6)^2}{6} + \frac{(6-6)^2}{6} + \frac{(5-8)^2}{8} + \frac{(6-9)^2}{9} + \frac{(9-9)^2}{9} + \frac{(15-12)^2}{12} = 4.375$$

Since $\chi^2 = 4.375$ does not exceed $9.21 = \chi^2_{.01}$ at degrees of freedom $(2-1) \times (3-1) = 2$, we cannot reject the null hypothesis of no relationship at a 99% level of confidence.

2.(6 pt) You selected simple random samples of adults from each of the following four states: Iowa, Massachusetts, New York, and Texas, and recorded how many members of each sample owned homes. The results are as follows:

(1) Iowa: Sample size = 50, 35 out of these 50 own homes.

(2) Massachusetts: Sample size = 50, 30 of these 50 own homes.

(3) New York: Sample size = 40, 21 out of these 40 own homes.

(4) Texas: Sample size = 60, 39 out of these 60 own homes.

Using the chi-square test, at a 99% level of confidence test the null hypothesis that equal proportions of adults in these four states own homes (that is, the null hypothesis $\pi_1 = \pi_2 = \pi_3 = \pi_4$).(Show work)

Answer: The given information can be expressed as a cross tabulation as follows:

State	No	Yes	Row Totals
Iowa	15	35	50
Mass	20	30	50
NY	19	21	40
Tx	21	39	60
Column Totals	75	125	$n = 200$

The expected frequencies are:

$$E_{11} = \frac{50 \times 75}{200} = 18.75 \quad E_{12} = \frac{50 \times 125}{200} = 31.25$$

$$E_{21} = \frac{50 \times 75}{200} = 18.75 \quad E_{22} = \frac{50 \times 125}{200} = 31.25$$

$$E_{31} = \frac{40 \times 75}{200} = 15 \quad E_{32} = \frac{40 \times 125}{200} = 25$$

$$E_{41} = \frac{60 \times 75}{200} = 22.5 \quad E_{42} = \frac{60 \times 125}{200} = 37.5$$

Since all expected frequencies are greater than 5, chi-square test is valid here.

$$\chi^2 = \frac{(15-18.75)^2}{18.75} + \frac{(35-31.25)^2}{31.25} + \frac{(20-18.75)^2}{18.75} + \frac{(30-31.25)^2}{31.35}$$

$$+\frac{(19-15)^2}{15}+\frac{(21-25)^2}{25}+\frac{(21-22.5)^2}{22.5}+(39-37.5)^2 37.5=3.2$$

Since $\chi^2 = 3.2$ does not exceed $11.34 = \chi^2_{.01}$ at degrees of freedom $(4-1)\times(2-1) = 3$, we cannot reject the null hypothesis of equal proportions at a 99% level of confidence.

3.(3+6 = 9 pt) Consider the regression model developed to examine how the demand of light beer at a store depends on brand name (which can be Amstel, Bud, Michelob, and Miller) and price:

$Y = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 + \beta_4 X + \beta_5 D_1 X + \beta_6 D_2 X + \beta_7 D_3 X + \epsilon,$

where:

Y = natural log of move, where move is the number of units (each unit is a six-pack of 12 oz bottles) of a brand of light beer sold at a store in a given week,

$D_1 = 1$ if the brand is Amstel, and $D_1 = 0$ if the brand is Bud, Michelob, or Miller,

$D_2 = 1$ if the brand is Bud, and $D_2 = 0$ if the brand is Amstel, Michelob, or Miller,

$D_3 = 1$ if the brand is Michelob, and $D_3 = 0$ if the brand is Amstel, Bud, or Miller,

X = natural logarithm of unit price.

As we discussed in class, if we run a regression with natural log of move as dependent variable and the natural log of price as independent variable for a single brand, the coefficient of natural log of price is the (own) price elasticity of demand of the brand. Thus here, the coefficient of X is the price elasticity of demand.

3.(a) Write the regression equation for each brand and express the price elasticity of demand of each brand in terms of the parameters of the regression model.

bb The regression equations for the four brands are given below:

Brand	Regression Equation	Price elasticity of demand
Amstel $D_1 = 1,\ D_2 = 0,\ D_3 = 0$ $D_1X = X,\ D_2X = 0,\ D_3X = 0$	$Y = (\beta_0+\beta_1)+(\beta_4+\beta_5)X+\epsilon$	$\beta_4+\beta_5$
Bud $D_1 = 0,\ D_2 = 1,\ D_3 = 0$ $D_1X = 0,\ D_2X = X,\ D_3X = 0$	$Y = (\beta_0+\beta_2)+(\beta_4+\beta_6)X+\epsilon$	$\beta_4+\beta_6$
Michelob $D_1 = 0,\ D_2 = 0,\ D_3 = 1$ $D_1X = 0,\ D_2X = 0,\ D_3X = X$	$Y = (\beta_0+\beta_3)+(\beta_4+\beta_7)X+\epsilon$	$\beta_4+\beta_7$
Miller $D_1 = 0,\ D_2 = 0,\ D_3 = 0$ $D_1X = 0,\ D_2X = 0,\ D_3X = 0$	$Y = \beta_0+\beta_4X+\epsilon$	β_4

For each brand, the price elasticity of demand is the coefficient of X.

3.(b) Express each of the following hypotheses in terms of the parameters of the regression model:

3.(b)(i) The price elasticity of demand is same for Amstel, Miller and Michelob.

Answer: $\beta_4 + \beta_5 = \beta_4 + \beta_7 = \beta_4 \longrightarrow \beta_5 = \beta_7 = 0$

3.(b)(ii) The price elasticity of demand is same for Bud and Michelob.

Answer: $\beta_4 + \beta_6 = \beta_4 + \beta_7 \longrightarrow \beta_6 - \beta_7 = 0$

3.(b)(iii) The regression line is same (that is, intercept and slope of X are both same) for Amstel and Miller.

Answer: $\beta_0 + \beta_1 = \beta_0 \longrightarrow \beta_1 = 0$, and $\beta_4 + \beta_5 = \beta_4 \longrightarrow \beta_5 = 0$. Summarizing, $\beta_1 = \beta_5 = 0$.

4.(3+3= 6 pt) Consider the following regression model:

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$,

where: Y = per capita dollar revenue in a sales territory in a given quarter;

X_1 = unit price;

X_2 = per capita advertising expenditure;

X_3 = per capita sales promotion expenditure;

X_4 = per capita personal selling expenditure;

ϵ is the random error defined as usual.

You collected data from a simple random sample of 45 territories and estimated several regression models. The results are as follows:

Regression #	Dependent Variable	Independent Variable(s)	R^2
1	Y	X_1, X_2, X_3, X_4	0.70
2	Y	X_1	0.30
3	Y	X_2	0.35
4	Y	X_3	0.20
5	Y	X_4	0.25
6	Y	X_1, X_2	0.40
7	Y	X_3, X_4	0.32
8	Y	X_1, X_3	0.55
9	Y	X_2, X_4	0.50
10	Y	X_1, X_2, X_3	0.60
11	Y	X_1, X_2, X_4	0.66
12	Y	X_1, X_3, X_4	0.65
13	Y	X_2, X_3, X_4	0.68

At a 99% level of confidence, test each of the following null hypotheses.

4.(a) $H_0 : \beta_1 = 0$

Answer: $n = 45$, $m = 4$, $k = 1$, $n - m - 1 = 45 - 4 - 1 = 40$

Full model: Model 1, $R^2_{full} = .70$

Restricted model: Model 13, $R^2 restricted = .68$

$F = (\frac{.70 - .68}{1 - .70}) \times (\frac{45 - 4 - 1}{1}) = 2.667$. Since $F = 2.667$ does not exceed $7.31 = F_{.01}(1, 40)$, we cannot reject reject H_0 at a 99% level of confidence.

4.(b) $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$

Answer: $n = 45$, $m = 4$, $k = 4$, $n - m - 1 = 45 - 4 - 1 = 40$

Full model: Model 1, $R^2_{full} = .70$

Restricted model: Naive model, $R^2 restricted = 0$

$F = (\frac{.70 - 0}{1 - .70}) \times (\frac{45 - 4 - 1}{4}) = 23.33 > 3.83 = F_{.01}(4, 40)$

Hence we reject H_0 at a 99% level of confidence.

5.($2 + 6 = 8$ pt) Consider the binary variable Y defined for an individual who has an account at a major bank. Y is 1 if the person took a loan from the bank in the last five years, and $Y = 0$ if not. A logit model is developed to examine how Y depends on the following independent variables:

- X = annual income of the person in units of \$1000. (For example, if income is \$80,000, then $X = 80$.)
- $D_1 = 1$ if the person has a graduate degree, and $D_1 = 0$ if not.
- $D_2 = 1$ if the person owns a home, and $D_2 = 0$ if not.

A standard logit model is used where $P(Y = 1|I) = \frac{1}{1 + e^{-I}}$ where the indicator function I is given by:

$I = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 X + \beta_4 D_1 X + \beta_5 D_2 X$

5.(a) State the following null hypothesis in terms of the parameters of the model: If an account holder has a graduate but does not own a home, then the probability of taking a loan does not depend on income.

Answer: If an account holder has a graduate degree but does not own a home, then $D_1 = 1$ and $D_2 = 0$, that is,

$I = \beta_0 + \beta_1 + (\beta_3 + \beta_4)X$

The hypothesis means I is same for this person for all values of X, that is, $\beta_3 + \beta_4 = 0$.

5.(b) Suppose the researcher obtained a random sample of account holders at the bank and estimated several logit models. Results are given below:

Model	Dependent variable	Independent variable(s)	$-2\ln L$
1	Y	D_1, D_2, X, D_1X, D_2X	480
2	Y	D_1, D_2	537
3	Y	X	528
4	Y	D_1, D_2, X	504
5	Y	D_1, X, D_1X	534
6	Y	D_2, X, D_2X	484

For the naive model, $-2\ln L = 570$.

At a 99% level of confidence, test each of the following the null hypotheses:

5(b)(i) $H_0 : \beta_2 = \beta_5 = 0$

Answer: $k = 2$

Full model: Model 1, $-2\ln L_{full} = 480$

Restricted model: Model 5, $-2\ln L_{restricted} = 534$

$\chi^2 = (-2\ln L_{restricted}) - (-2\ln L_{full}) = 534 - 480 = 54 > 9.21 = \chi^2_{.01}(2) \longrightarrow$ reject H_0 at a 99% level of confidence.

5(b)(ii) $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$.

Answer: $k = 5$

Full model: Model 1, $-2\ln L_{full} = 480$

Restricted model: Naive model, $-2\ln L_{restricted} = 570$

$\chi^2 = (-2\ln L_{restricted}) - (-2\ln L_{full}) = 570 - 480 = 90 > 15.09 = \chi^2_{.01}(5) \longrightarrow$ reject H_0 at a 99% level of confidence.

6.(5 pt) Consider the binary variable Y which is 1 if a person defaults on a loan, and 0 if she does not. The probability that $Y = 1$ is given by the LOGIT model with indicator function

$I = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 D$,

where X_1 = annual income (in units of \$1000), X_2 = number of years the person has been employed, X_3 = the person's rent or mortgage as fraction of income, and $D = 1$ if the person is married and zero if not. You also know that $\beta_0 = -2.1$, $\beta_1 = -.002$, $\beta_2 = -.02$, $\beta_3 = 1.6$, and $\beta_4 = -.6$.

A person who is married, has annual income \$100,000, has been employed for 10 years, and has a rent that is 30% of income has applied for a loan from a banker. If the banker does not approve the loan and Y is really zero, the banker has an opportunity loss of \$9000. If the banker approves the loan and the person defaults (that is, $Y = 1$), the banker loses \$100,000.

For this person, compute $P(Y = 1)$ and the odds ratio. Based on the criterion of minimizing expected loss, should the banker give this person the loan? (Show work)

Answer: For this person, $X_1 = 100$, $X_2 = 10$, $X_3 = .3$, and $D = 1$. Hence,

$I = -2.1 - (.002 \times 100) - (.02 \times 10) + (1.6 \times .3) - .6 = -2.62$

$P(Y = 1) = \dfrac{1}{1 + e^{2.62}} = .0679$

$P(Y = 0) = 1 - P(Y = 1) = 1 - .0679 = .9321$

Odds ratio $= \dfrac{.0679}{.9321} = .0728$

(Alternatively, odds ratio $= e^I = e^{-2.62} = .0728$.)

$C(1|0) = 9000$, $C(0|1) = 100,000$, $\dfrac{C(1|0)}{C(0|1)} = .09$

Since odds ratio is less than $\dfrac{C(1|0)}{C(0|1)}$, we assign this applicant to $Y = 0$ (not default) and approve the loan.

Alternative method 1: $I = -2.62 < -2.408 = \ln(.09) = \ln(\dfrac{C(1|0)}{C(0|1)}$

Hence, we assign this case to $Y = 0$ (not default) and approve the loan.

Alternative method 2: Expected loss if loan is approved

$= C(0|1) * P(Y = 1) + 0 * P(Y = 0) = 100,000 * .0679 = 6790$

Expected loss if loan is not approved

$= 0 * P(Y = 1) + C(1|0) * P(Y = 0) = 9000 * .9321 = 8389$

Since $6790 < 8389$, approve the loan.

7.(5 pt) Conjoint analysis is done to determine how a person's evaluation of an electric car depends on the levels of two attributes:

- X_1: The number of highway miles one can drive the car after a full charge. Range: 200-400 miles.
- X_2: The time it takes to recharge the car after a full discharge. Range: 6-24 hours.

It is assumed that a person's rating of an electric car on a 0-100 (very poor to excellent) scale can be expressed as:

$$U(X_1, X_2) \quad = \quad U_0 + U_1(X_1) + U_2(X_2),$$

where $U_1(200) = 0$, $U_2(6) = 0$, and a higher score means the person likes the product better.

Suppose, for a given person:

$U_0 = 40$		
$U_1(200) = 0$	$U_1(300) = 40$	$U_1(400) = 60$
$U_2(6) = 0$	$U_2(12) = -10$	$U_2(24) = -40$

Which of the following two electric cars will this person prefer?

- Car 1: $X_1 = 360$, $X_2 = 21$
- Car 2: $X_1 = 240$, $X_2 = 9$

(Show all numbers used in computations and briefly explain how you computed the numbers)

Answer: $U_1(240) = U_1(200) + [\{\frac{240-200}{300-200}\} \times \{U_1(300) - U_1(200)\}]$

$= 0 + [(\frac{40}{100}) \times \{40 - 0\}] = 16$

$U_1(360) = U_1(300) + [\{\frac{360-300}{400-300}\} \times \{U_1(400) - U_1(300)\}]$

$= 40 + [(\frac{60}{100}) \times \{60 - 40\}] = 52$

$U_2(9) = U_2(6) + [(\frac{9-6}{12-6}) \times \{U_2(12) - U_2(6)\}] = 0 + [(\frac{3}{6}) \times (-10 - 0)] = -5$

$U_2(21) = U_2(12) + [(\frac{21-12}{24-12}) \times \{U_2(24) - U_2(12)\}]$

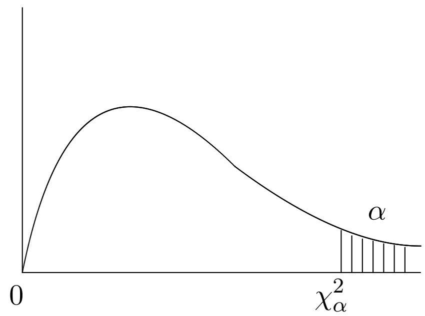
$= -10 + [(\frac{9}{12}) \times \{(-40) - (-10)\} = -32.5$

Car 1: Score $= U_0 + U_1(360) + U_2(21) = 40 + 52 - 32.5 = 59.5$

Car 2: Score $= U_0 + U_1(240) + U_2(9) = 40 + 16 - 5 = 51$

Since $59.5 > 51$, this person prefers Car 1.

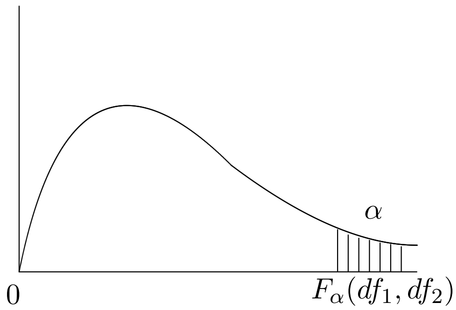
Table 2. Chi-Square (χ^2) Distribution



	α				
df	**.10**	**.05**	**.025**	**.01**	**.005**
1	2.71	3.84	5.02	6.63	7.88
2	4.61	5.99	7.38	9.21	10.60
3	6.25	7.81	9.35	11.34	12.84
4	7.78	9.49	11.14	13.28	14.86
5	9.24	11.07	12.83	15.09	16.75
6	10.64	12.59	14.45	16.81	18.55
7	12.02	14.07	16.01	18.48	20.28
8	13.36	15.51	17.53	20.09	21.95
9	14.68	16.92	19.02	21.67	23.59
10	15.99	18.31	20.48	23.21	25.19
11	17.28	19.68	21.92	24.73	26.76
12	18.55	21.03	23.34	26.22	28.30
13	19.81	22.36	24.74	27.69	29.82
14	21.06	23.68	26.12	29.14	31.32
15	22.31	25.00	27.49	30.58	32.80
16	23.54	26.30	28.85	32.00	34.27
17	24.77	27.59	30.19	33.41	35.72
18	25.99	28.87	31.53	34.81	37.16
19	27.20	30.14	32.85	36.19	38.58
20	28.41	31.41	34.17	37.57	40.00
21	29.62	32.67	35.48	38.93	41.40
22	30.81	33.92	36.78	40.29	42.80
23	32.01	35.17	38.08	41.64	44.18
24	33.20	36.42	39.36	42.98	45.56
25	34.38	37.65	40.65	44.31	46.93
26	35.56	38.89	41.92	45.64	48.29
27	36.74	40.11	43.19	46.96	49.65
28	37.92	41.34	44.46	48.28	50.99
29	39.09	42.56	45.72	49.59	52.34
30	40.26	43.77	46.98	50.89	53.67

Table 4b. F Distribution ($\alpha = .01$)

(df_1 **= Degrees of Freedom of Numerator,** df_2 **= Degrees of Freedom of Denominator)**



	df_1											
df_2	1	2	3	4	5	6	7	8	9	10	12	15
1	4052	4999	5403	5624	5764	5859	5928	5981	6022	6056	6107	6157
2	98.50	99.00	99.16	99.25	99.30	99.33	99.36	99.38	99.39	99.40	99.42	99.43
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	27.05	26.87
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	14.37	14.20
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.89	9.72
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.99	2.85
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	2.96	2.81
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06	2.93	2.78
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.90	2.75
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00	2.87	2.73
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04