Student Name: ______________________ ID: ______________________

# McGill University

# Faculty of Science

# Department of Mathematics and Statistics

## Statistics Part A Comprehensive Exam

## Methodology Paper

Date: Friday, May 16, 2014

Time: 13:00 – 17:00

## Instructions

- Answer only **two** questions out of Section L. If you answer more than two questions, then only the **FIRST TWO questions will be marked.**
- Answer only **two** questions out of Section G. If you answer more than two questions, then only the **FIRST TWO questions will be marked.**

| Questions | Marks |
|---|---|
| L1 | |
| L2 | |
| L3 | |
| G1 | |
| G2 | |
| G3 | |

This exam comprises the cover page and 16 pages of questions.

**Section L (Linear Regression Models)**
**Answer only two questions out of L1–L3**

**L1.**

(a) (**The use of indicator variables in linear regression analysis**):

Consider a simple linear regression model that is used to study the relationship between a response variable $y$ and a covariate $x$. The analysis is based on $n$ independent observations

$$(x_i, y_i) \ , \ \ i = 1, 2, \ldots, n.$$

Suppose these $n$ observations are divided into $M$ groups each having $n_m$ observations, such that $\sum_{m=1}^{M} n_m = n$. Assume that the groups are independent from each other. The most general simple linear regression model that can be used to analyze such data is

$$y = \beta_{0m} + \beta_{1m} x + \varepsilon \ , \ \ m = 1, 2, \ldots, M$$

where $\varepsilon \sim N(0, \sigma^2)$. That is, one can fit $M$ separate (different) simple linear regression models to the data, corresponding to the $M$ independent groups of observations.

Using indicator variables, **explain in detail**, how we can formally test the null hypothesis

$$H_0 : \beta_{11} = \beta_{12} = \ldots = \beta_{1M}.$$

That is, we want to test whether we should fit $M$ parallel lines to the data, corresponding to the $M$ groups. **(15 marks)**

(b) (**One-way analysis of variance models**):

Sociologists often conduct experiments to investigate the relationship between socioeconomic status and college performance. Socioeconomic status is generally partitioned into three classes: lower, middle and upper class. Consider the problem of comparing the mean grade point averages of the college freshmen between the three classes. The grade point averages for random samples of four college freshmen associated with each of the three socioeconomic classes were selected from a university's files at the end of the academic year. The data are given in the table on the next page.

| Grade Point Averages | | |
|---|---|---|
| **Lower class** | **Middle class** | **Upper class** |
| 2.87 | 3.23 | 2.25 |
| 2.16 | 3.45 | 3.13 |
| 3.14 | 2.78 | 2.44 |
| 2.51 | 3.77 | 2.54 |

Denote

$\mu_0$: grand mean of the grade point average for a college freshman.

$\tau_i$: the effect of the $i$-th socioeconomic class on the grade point average, for $i = 1, 2, 3$.

$\mu_i = \mu_0 + \tau_i$: mean of the grade point average for a college freshman in the $i$-th socioeconomic class, for $i = 1, 2, 3$.

In the above notation $i = 1, 2, 3$, are corresponding to the three socioeconomic classes mentioned in the table, respectively.

(i) Write down a **one-way analysis of variance** model that is used to analyze such data. State what assumptions one typically makes when using this model. (**8 marks**)

(ii) Using indicator variables, write down a *multiple linear regression model* equivalent to the one-way analysis of variance model in **Part(i)**. (**3 marks**)

(iii) Using the data in the above table, write down the design matrix $\boldsymbol{X}$ and the vector $\boldsymbol{Y}$ for the regression model in **Part (ii)**. Find the value of least-squares estimate of the vector of regression parameters $\boldsymbol{\beta}$ by solving the normal equations $(\boldsymbol{X}^\top \boldsymbol{X})\,\boldsymbol{\beta} = \boldsymbol{X}^\top \boldsymbol{Y}$ *directly*. (**3 marks**)

(iv) Using the least-squares estimates from **Part(iii)**, find the value of the least-squares estimates of the parameters $(\mu_0, \tau_1, \tau_2, \tau_3)$ of the model in **Part(i)**. (**3 marks**)

(v) A sociologist wants to test the null hypothesis

$$H_0 : \tau_1 = \tau_2 = \tau_3 = 0 \ \ (\text{NO CLASS EFFECT ON THE GRADE POINT AVERAGE})$$

in the analysis of variance model in **Part (i)**. Using the regression model in **Part (ii)**, ***derive*** an appropriate test statistics that is used to test $H_0$ at the significance level $\alpha$. (**8 marks**)

**L2.**

An inverter is an electrical device that converts direct current (DC) to alternating current (AC). The data in this question are on measurement of the transient points of an electronic inverter. A portion of the data is given in the following Table. There are $24$ observations.

Table 1:

| | $x_1$ | $x_2$ | $x_3$ | $x_4$ | Y |
|---|---|---|---|---|---|
| 1 | 3.00 | 3.00 | 3.00 | 3.00 | 0.79 |
| 2 | 3.00 | 6.00 | 6.00 | 6.00 | 1.71 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 23 | 2.00 | 3.00 | 8.00 | 6.00 | 1.51 |
| 24 | 3.00 | 3.00 | 8.00 | 8.00 | 0.75 |

The variables of interest are:

$Y$: Transient point (volts) of PMOS-NMOS inverters.
$x_1$: Width of the NMOS device.
$x_2$: Length of the NMOS device.
$x_3$: Width of the PMOS device.
$x_4$: Length of the PMOS device

**Question L2 is continued on the next page.**

(a) The following multiple linear regression model is used to study the relationship between $Y$ and the regressors:

$$\textbf{Model 1:}\ \ Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i\ ,\ i = 1, 2, \ldots, 24.$$

The R-output for **Model (1)** is given on pages 7 and 8. Use the significance level $5\%$ in your analysis.

(Provide concrete statistical reasoning for your answers.)

(i) The individual contribution of each covariate in the fitted model. (**2 marks**)

(ii) Overall significance and also performance of the fitted model. (**4 marks**)

(iii) Based on the two residual plots in Figure 1 on page 8, comment on the appropriateness of the fitted **Model 1**. Provide concrete reasoning whether you would suggest this model for the use in practice or not. (**8 marks**)

It has been suggested that a suitable model for these data is:

$$Y = \alpha_0 \left(x_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3} x_4^{\alpha_4}\right) \times \varepsilon^*.$$

(b) By using an appropriate transformation to the above model, write down its corresponding **multiple linear regression** model, and call it **Model 2**. What are the the new regressors and response variable in **Model 2**? (Write down the new variables as functions of the original variables.) What are the usual assumptions on the error term in **Model 2**? (**12 marks**)

(c) The new multiple linear regression **Model 2** discussed in **Part (b)** is fitted to the data. The **R** output is given on page 9. Based on the R output for **Model 2** and the two residual plots in Figure 2 on page 10, comment on the appropriateness of the fitted **Model 2**. Provide concrete reasoning whether you would suggest this model, compared to **Model 1**, for the use in practice or not. (**14 marks**)

**L3.**

Consider the multiple linear regression model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \ldots, \beta_k)$ is the unknown vector of regression parameters, $\boldsymbol{Y}$ is the $n \times 1$ dimensional vector of observations of the response variable $y$, $\boldsymbol{X}$ is the $n \times (k+1)$ dimensional design matrix; $\boldsymbol{\varepsilon}$ is the $n \times 1$ dimensional vector of errors $\varepsilon_i$, and $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{V})$, where $\boldsymbol{V}$ is a known matrix and $\sigma^2$ is unknown.

(a)

(i) Write down the ordinary least squares (OLS) estimator of $\boldsymbol{\beta}$. In one or two sentences, without any derivation, explain why the OLS estimator is not an appropriate estimator for $\boldsymbol{\beta}$ in the above model. **(6 marks)**

(ii) Derive the generalized least squares estimator of $\boldsymbol{\beta}$. What is the advantage of this estimator of $\boldsymbol{\beta}$ compared to the one in **Part (i)**? (No need to provide any mathematical proof for your answer.) **(6 marks)**

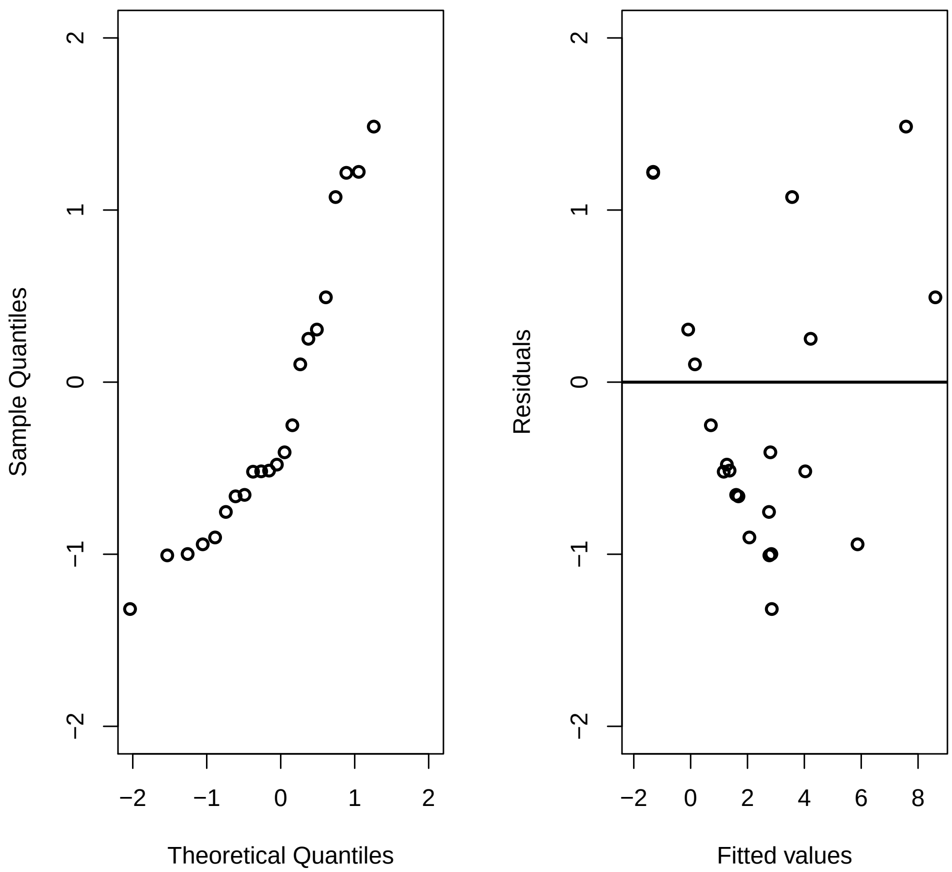
(iii) Derive the distribution of generalized least squares estimator of $\boldsymbol{\beta}$. Write down the analysis of variance (ANOVA) table, and explain how to formally test *the overall significance* of the model. You need to provide an appropriate test statistics to perform this test. **(8 marks)**

(b) Explain in detail what the *weighed least squares estimator* of the parameter vector $\boldsymbol{\beta}$ is in a multiple linear regression model. Write down the weighed least squares estimator of $\boldsymbol{\beta}$ in the form of a generalized least squares estimator discussed in **Part (a-ii)**. **(20 marks)**

## R-Code for Question L2.

```
> D1<-read.table("Q2_data.txt",header=T)
> attach(D1)
> fit1<-lm(y~.,data=D1)
> summary(fit1)

Call:
lm(formula = y ~ ., data = D1)

Residuals:
    Min      1Q  Median      3Q     Max
-1.4894 -0.9324 -0.6098  0.7224  3.3659

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.52430    1.02317   1.490 0.152692
x1          -0.30606    0.07736  -3.956 0.000847 ***
x2           0.37439    0.05820   6.433 3.63e-06 ***
x3           0.44957    0.12354   3.639 0.001746 **
x4          -0.46557    0.13750  -3.386 0.003102 **
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1

Residual standard error: 1.445 on 19 degrees of freedom
Multiple R-squared: 0.8044,     Adjusted R-squared: 0.7632
F-statistic: 19.53 on 4 and 19 DF,  p-value: 1.604e-06

> anova(fit1)
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
x1         1  11.989  11.989  5.7385  0.027056 *
x2         1 111.745 111.745 53.4885 6.185e-07 ***
x3         1  15.523  15.523  7.4304  0.013418 *
x4         1  23.951  23.951 11.4645  0.003102 **
Residuals 19  39.694   2.089
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1
```

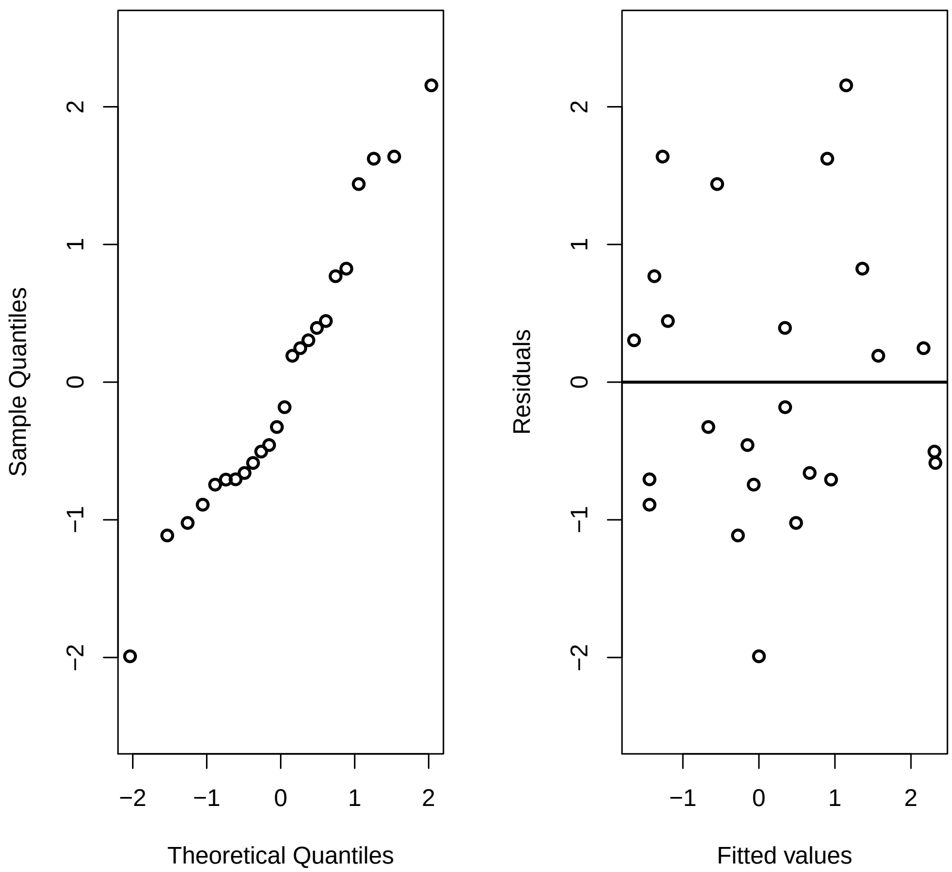
Figure 1: Residuals plots for **Model 1**.



```
> fit2<-lm(new_y~.,data=D1)
> summary(fit2)

Call:
lm(formula = new_y ~ ., data = D1)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.36567 -0.13258 -0.04697  0.10125  0.40247 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.07485    0.21386   -0.35     0.73    
new_x1     -1.24866    0.07939  -15.73 2.38e-12 ***
new_x2      1.58623    0.07811   20.31 2.41e-14 ***
new_x3      0.93365    0.10515    8.88 3.44e-08 ***
new_x4     -1.34004    0.11749  -11.41 6.08e-10 ***
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1

Residual standard error: 0.2086 on 19 degrees of freedom
Multiple R-squared:  0.9772,    Adjusted R-squared:  0.9724 
F-statistic: 203.7 on 4 and 19 DF,  p-value: 2.57e-15

> anova(fit2)
Analysis of Variance Table

Response: new_y
          Df  Sum Sq Mean Sq F value    Pr(>F)    
new_x1     1  3.2553  3.2553  74.832 5.149e-08 ***
new_x2     1 26.0375 26.0375 598.539 7.940e-16 ***
new_x3   1  0.4960  0.4960  11.402  0.003167 **
new_x4     1  5.6591  5.6591 130.089 6.080e-10 ***
Residuals 19  0.8265  0.0435                      
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1
```

**Note:** The response and covariates in **Model 2** are called “new” in the output.

Figure 2: Residuals plots for **Model 2**.



**Section G (Generalized Linear Models)**
**Answer only two questions out of G1–G3**

**G1.** The table below summarizes some data collected on traffic accidents and seat belt usage collected in the state of Florida in 1988:

| Safety Equipment In Use ($S$) | Whether Ejected ($E$) | Injury ($I$) Nonfatal | Injury ($I$) Fatal |
|---|---|---|---|
| Seat belt | Yes | 1105 | 14 |
| | No | 411,111 | 483 |
| None | Yes | 4624 | 497 |
| | No | 157,342 | 1008 |

Table 2: *Source:* Florida Department of Highway Safety and Motor Vehicles.

**(a)** Describe briefly why one cannot test for goodness of fit of the three-way interaction model, $S.E.I$, *in 3 sentences or fewer*. (**2 marks**)

**(b)** Consider the *homogeneous association* or *no three-way interaction* Poisson log-linear model, $S.E + E.I + S.I$. Describe, in words, what this model assumes about the conditional independence structure of the data. (**2 marks**)

**(c)** Identify the model (`gmodel1a`, `gmodel1b`, or `gmodel1c`) in the R output on pages 14 and 15 which fits the model of homogeneous association to the Florida seat belt data. Assess the goodness of fit of the model you've identified using a significance level $\alpha = 0.01$. (**8 marks**)

**(d)** Identify the model (`gmodel1a`, `gmodel1b`, or `gmodel1c`) in the R output on pages 14 and 15 which fits the following model: $S.E + E.I$. Assess the goodness of fit of the model you've identified using a significance level $\alpha = 0.01$. (**8 marks**)

**Question G1 is continued on the next page.**

**(e)** Consider now the two models that you selected in parts (c) and (d) on the previous page. Which is the more appropriate model? Explain your reasoning. **(6 marks)**

**(f)** Another public safety researcher modeled the data in Table 2 using the following logistic regression model:

$$\log\left(\frac{Pr(I = \text{Nonfatal}|S, E)}{Pr(I = \text{Fatal}|S, E)}\right) = \beta_0 + \beta_S I(S = \text{Yes})$$

Determine which of the three Poisson loglinear models on pages 14 and 15 yields equivalent inference to this logistic regression model and prove that the models are equivalent. **(8 marks)**

**(g)** Compute the maximum likelihood parameter estimates for $\hat{\beta}_0$ and $\hat{\beta}_S$ for the logistic regression model in part (f) using the equivalent Poisson loglinear model that you identified in part (f). **(6 marks)**

**G2.**

(a) Suppose in an independent, two-sample problem that $Y_i$ is Poisson with $g(\mu_i) = \alpha + \beta x_i$, where $x_i = 1$ for $i = 1, .., n_A$ for the observations from group A and $x_i = 0$ for $i = n_A + 1, ..., n$ for the observations from group B. Show that for any link function $g$, the score equations imply that fitted means $\hat{\mu}_A$ and $\hat{\mu}_B$ equal the sample means. **(10 marks)**

(b) In a GLM, suppose that $Var(Y) = v(\mu)$ for $\mu = E(Y)$. Show that the link $g$ satisfying $g'(\mu) = [v(\mu)]^{-1/2}$ has the same weight matrix $\mathbf{W}^{(t)}$ at each cycle of the Fisher scoring algorithm. Show that this link function for a Poisson GLM is $g(\mu) = 2\sqrt{\mu}$.

**(10 marks)**

(c) Assume that $Y$ is an ordinal response variable and $x$ is a continuous predictor. Show that for the cumulative logit model,

$$\text{logit}[Pr(Y \leqslant j)] = \alpha_j + \beta x$$

cumulative probabilities may be misordered for some x values. **(10 marks)**

(d) Assume that $X$ and $Y$ are both ordinal categorical variables and $Z$ is binary variable. Explain how to test conditional independence of $Y$ and $X$ given $Z$ by generalizing the cumulative logit model to allow a different trend in each partial table.

**(10 marks)**

**G3.** An often used data set in statistics courses contains observations of female horseshoe crabs. For this problem, $Y_i$ is the binary response variable of interest is whether the $i-$th horseshoe crab has at least one other male living outside of her nest (a satellite) or not. The only covariate of interest ($X_i$) for this problem is the weight (in kg) of the female crab. The goal of the analysis is to characterize the relationship between weight of the female horseshoe crabs and the presence of at least one male satellite crab.

(a) A marine biologist collaborator tells you that their past experience indicates that the probability of the presence of at least one satellite crab varies linearly with the weight of the female crab. Your collaborator suggests that one should use simple linear regression with $Y$ as the response and $X$ as the sole covariate. Give **at least two** reasons, without referring to the output, why a simple linear regression model would not necessarily be appropriate here. **4 marks**

(b) You suggest that a GLM with a logistic link would be a better choice. Your collaborator is concerned that this model would not respect the assumption that the probability of a satellite was approximately linear in the weight. Again without referring to the R output, explain why using a generalized linear model with a logistic link would still be reasonable if the association were truly linear over the support of $X$. *Hint:* Characterize how the logistic link function varies with $X$.

**8 marks**

(c) For this question refer to the R output on page 16. Test for the presence of an association between the weight of the horseshoe crab and the presence of a satellite male at $\alpha = 0.05$ using the logistic regression model. Interpret the value of the coefficient for weight in the context of the real data set.

**8 marks**

(d) Again, using the R output on page 16, assess the claim made by your collaborator about the linear relationship between the probability of response and the weight of the crab. Do you think that using simple linear regression for this dataset is reasonable? Explain why or why not.

**10 marks**

(e) Write the objective function that the estimated parameters for the simple linear regression model minimize in terms of the estimated probabilities of response, $\hat{p}_i^{LM}$. Write the objective function that the estimated parameters for the logistic regression model minimize in terms of the estimated probabilities of response, $\hat{p}_i^{GLM}$. Comment on the similarities and differences between the two objective functions.

**10 marks**

## Code and Output for Question G1.

```
> summary(gmodel1a)

Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
(Intercept)           6.92251    0.03110  222.56   <2e-16 ***
SSeat belt           -0.75682    0.05394  -14.03   <2e-16 ***
EYes                 -0.72784    0.05345  -13.62   <2e-16 ***
INonfatal             5.04362    0.03120  161.65   <2e-16 ***
SSeat belt:EYes      -2.39964    0.03334  -71.97   <2e-16 ***
SSeat belt:INonfatal  1.71732    0.05402   31.79   <2e-16 ***
EYes:INonfatal       -2.79779    0.05526  -50.63   <2e-16 ***
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1.6249e+06  on 7  degrees of freedom
Residual deviance: 2.8540e+00  on 1  degrees of freedom
AIC: 93.853

Number of Fisher Scoring iterations: 3
```

```
> summary(gmodel1b)

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)       6.026472   0.025986 231.916   <2e-16 ***
SSeat belt        0.955230   0.002957 323.024   <2e-16 ***
EYes              0.012267   0.051645   0.238    0.812
INonfatal         5.943472   0.025932 229.198   <2e-16 ***
SSeat belt:EYes -2.476144   0.033131 -74.738   <2e-16 ***
EYes:INonfatal  -3.526545   0.052952 -66.599   <2e-16 ***
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1624865.3  on 7  degrees of freedom
Residual deviance:    1144.6  on 2  degrees of freedom
AIC: 1233.6

Number of Fisher Scoring iterations: 5

> summary(gmodel1c)

Coefficients:
                      Estimate Std. Error z value Pr(>|z|)
(Intercept)            7.28472    0.02578  282.57   <2e-16 ***
SSeat belt            -1.07885    0.05174  -20.85   <2e-16 ***
EYes                  -3.43146    0.01420 -241.68   <2e-16 ***
INonfatal              4.67859    0.02590  180.67   <2e-16 ***
SSeat belt:EYes       -2.47614    0.03313  -74.74   <2e-16 ***
SSeat belt:INonfatal  2.04212    0.05182   39.41   <2e-16 ***
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 1624865.3  on 7  degrees of freedom
Residual deviance:    1680.4  on 2  degrees of freedom
AIC: 1769.4

Number of Fisher Scoring iterations: 5
```

## Code and Output for Question G3.

```
> gmodel3a = glm(Y~X,data=crabs)
> summary(gmodel3a)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8878  -0.4683   0.1606   0.3704   0.6689

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.14487    0.14715  -0.984    0.326
X            0.32270    0.05876   5.492 1.42e-07 ***
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1

(Dispersion parameter for gaussian family taken to be 0.1977574)

    Null deviance: 39.780  on 172  degrees of freedom
Residual deviance: 33.817  on 171  degrees of freedom
AIC: 214.56

Number of Fisher Scoring iterations: 2

> gmodel3b = glm(Y~X,data=crabs,family=binomial(logit))
> summary(gmodel3b)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1108  -1.0749   0.5426   0.9122   1.6285

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.6947     0.8802  -4.198 2.70e-05 ***
X             1.8151     0.3767   4.819 1.45e-06 ***
---
Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1     1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 225.76  on 172  degrees of freedom
Residual deviance: 195.74  on 171  degrees of freedom
AIC: 199.74

Number of Fisher Scoring iterations: 4
```

# Table of the Chi-squared distribution

Entries in table are $\chi^2_\alpha(\nu)$: the $\alpha$ tail quantile of Chi-squared($\nu$) distribution
$\alpha$ given in columns, $\nu$ given in rows.

| | Left-tail | | | | | Right-tail | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $\nu$ | 0.99500 | 0.99000 | 0.97500 | 0.95000 | 0.90000 | 0.10000 | 0.05000 | 0.02500 | 0.01000 | 0.00500 |
| 1 | 0.00004 | 0.00016 | 0.00098 | 0.00393 | 0.01579 | 2.70554 | 3.84146 | 5.02389 | 6.63490 | 7.87944 |
| 2 | 0.01003 | 0.02010 | 0.05064 | 0.10259 | 0.21072 | 4.60517 | 5.99146 | 7.37776 | 9.21034 | 10.59663 |
| 3 | 0.07172 | 0.11483 | 0.21580 | 0.35185 | 0.58437 | 6.25139 | 7.81473 | 9.34840 | 11.34487 | 12.83816 |
| 4 | 0.20699 | 0.29711 | 0.48442 | 0.71072 | 1.06362 | 7.77944 | 9.48773 | 11.14329 | 13.27670 | 14.86026 |
| 5 | 0.41174 | 0.55430 | 0.83121 | 1.14548 | 1.61031 | 9.23636 | 11.07050 | 12.83250 | 15.08627 | 16.74960 |
| 6 | 0.67573 | 0.87209 | 1.23734 | 1.63538 | 2.20413 | 10.64464 | 12.59159 | 14.44938 | 16.81189 | 18.54758 |
| 7 | 0.98926 | 1.23904 | 1.68987 | 2.16735 | 2.83311 | 12.01704 | 14.06714 | 16.01276 | 18.47531 | 20.27774 |
| 8 | 1.34441 | 1.64650 | 2.17973 | 2.73264 | 3.48954 | 13.36157 | 15.50731 | 17.53455 | 20.09024 | 21.95495 |
| 9 | 1.73493 | 2.08790 | 2.70039 | 3.32511 | 4.16816 | 14.68366 | 16.91898 | 19.02277 | 21.66599 | 23.58935 |
| 10 | 2.15586 | 2.55821 | 3.24697 | 3.94030 | 4.86518 | 15.98718 | 18.30704 | 20.48318 | 23.20925 | 25.18818 |
| 11 | 2.60322 | 3.05348 | 3.81575 | 4.57481 | 5.57778 | 17.27501 | 19.67514 | 21.92005 | 24.72497 | 26.75685 |
| 12 | 3.07382 | 3.57057 | 4.40379 | 5.22603 | 6.30380 | 18.54935 | 21.02607 | 23.33666 | 26.21697 | 28.29952 |
| 13 | 3.56503 | 4.10692 | 5.00875 | 5.89186 | 7.04150 | 19.81193 | 22.36203 | 24.73560 | 27.68825 | 29.81947 |
| 14 | 4.07467 | 4.66043 | 5.62873 | 6.57063 | 7.78953 | 21.06414 | 23.68479 | 26.11895 | 29.14124 | 31.31935 |
| 15 | 4.60092 | 5.22935 | 6.26214 | 7.26094 | 8.54676 | 22.30713 | 24.99579 | 27.48839 | 30.57791 | 32.80132 |
| 16 | 5.14221 | 5.81221 | 6.90766 | 7.96165 | 9.31224 | 23.54183 | 26.29623 | 28.84535 | 31.99993 | 34.26719 |
| 17 | 5.69722 | 6.40776 | 7.56419 | 8.67176 | 10.08519 | 24.76904 | 27.58711 | 30.19101 | 33.40866 | 35.71847 |
| 18 | 6.26480 | 7.01491 | 8.23075 | 9.39046 | 10.86494 | 25.98942 | 28.86930 | 31.52638 | 34.80531 | 37.15645 |
| 19 | 6.84397 | 7.63273 | 8.90652 | 10.11701 | 11.65091 | 27.20357 | 30.14353 | 32.85233 | 36.19087 | 38.58226 |
| 20 | 7.43384 | 8.26040 | 9.59078 | 10.85081 | 12.44261 | 28.41198 | 31.41043 | 34.16961 | 37.56623 | 39.99685 |
| 21 | 8.03365 | 8.89720 | 10.28290 | 11.59131 | 13.23960 | 29.61509 | 32.67057 | 35.47888 | 38.93217 | 41.40106 |
| 22 | 8.64272 | 9.54249 | 10.98232 | 12.33801 | 14.04149 | 30.81328 | 33.92444 | 36.78071 | 40.28936 | 42.79565 |
| 23 | 9.26042 | 10.19572 | 11.68855 | 13.09051 | 14.84796 | 32.00690 | 35.17246 | 38.07563 | 41.63840 | 44.18128 |
| 24 | 9.88623 | 10.85636 | 12.40115 | 13.84843 | 15.65868 | 33.19624 | 36.41503 | 39.36408 | 42.97982 | 45.55851 |
| 25 | 10.51965 | 11.52398 | 13.11972 | 14.61141 | 16.47341 | 34.38159 | 37.65248 | 40.64647 | 44.31410 | 46.92789 |
| 26 | 11.16024 | 12.19815 | 13.84390 | 15.37916 | 17.29188 | 35.56317 | 38.88514 | 41.92317 | 45.64168 | 48.28988 |
| 27 | 11.80759 | 12.87850 | 14.57338 | 16.15140 | 18.11390 | 36.74122 | 40.11327 | 43.19451 | 46.96294 | 49.64492 |
| 28 | 12.46134 | 13.56471 | 15.30786 | 16.92788 | 18.93924 | 37.91592 | 41.33714 | 44.46079 | 48.27824 | 50.99338 |
| 29 | 13.12115 | 14.25645 | 16.04707 | 17.70837 | 19.76774 | 39.08747 | 42.55697 | 45.72229 | 49.58788 | 52.33562 |
| 30 | 13.78672 | 14.95346 | 16.79077 | 18.49266 | 20.59923 | 40.25602 | 43.77297 | 46.97924 | 50.89218 | 53.67196 |
| 40 | 20.70654 | 22.16426 | 24.43304 | 26.50930 | 29.05052 | 51.80506 | 55.75848 | 59.34171 | 63.69074 | 66.76596 |
| 50 | 27.99075 | 29.70668 | 32.35736 | 34.76425 | 37.68865 | 63.16712 | 67.50481 | 71.42020 | 76.15389 | 79.48998 |
| 60 | 35.53449 | 37.48485 | 40.48175 | 43.18796 | 46.45889 | 74.39701 | 79.08194 | 83.29767 | 88.37942 | 91.95170 |
| 70 | 43.27518 | 45.44172 | 48.75756 | 51.73928 | 55.32894 | 85.52704 | 90.53123 | 95.02318 | 100.42518 | 104.21490 |
| 80 | 51.17193 | 53.54008 | 57.15317 | 60.39148 | 64.27784 | 96.57820 | 101.87947 | 106.62857 | 112.32879 | 116.32106 |
| 90 | 59.19630 | 61.75408 | 65.64662 | 69.12603 | 73.29109 | 107.56501 | 113.14527 | 118.13589 | 124.11632 | 128.29894 |
| 100 | 67.32756 | 70.06489 | 74.22193 | 77.92947 | 82.35814 | 118.49800 | 124.34211 | 129.56120 | 135.80672 | 140.16949 |

# Table of the F-distribution distribution

Entries in table are $F_{1-\alpha}(\nu_1, \nu_2)$: the $\alpha$ tail quantile of the $F_{\nu_1,\nu_2}$ distribution

| $\nu_2 \backslash \nu_l$ | $1-\alpha$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 10 | 12 | 15 | 20 | 30 | 50 | $\infty$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.900 | 49.5 | 53.6 | 55.8 | 57.2 | 58.2 | 59.1 | 59.7 | 60.5 | 61.0 | 61.5 | 62.0 | 62.6 | 63.0 | 63.3 |
| | 0.950 | 199. | 216. | 225. | 230. | 234. | 237. | 239. | 242. | 244. | 246. | 248. | 250. | 252. | 254. |
| | 0.975 | 800. | 864. | 900. | 922. | 937. | 948. | 957. | 969. | 977. | 985. | 993. | | | |
| | 0.990 | | | | | | | | | | | | | | |
| | 0.999 | | | | | | | | | | | | | | |
| 2 | 0.900 | 9.00 | 9.16 | 9.24 | 9.29 | 9.33 | 9.35 | 9.37 | 9.39 | 9.41 | 9.43 | 9.44 | 9.46 | 9.47 | 9.49 |
| | 0.950 | 19.0 | 19.2 | 19.2 | 19.3 | 19.3 | 19.4 | 19.4 | 19.4 | 19.4 | 19.4 | 19.4 | 19.5 | 19.5 | 19.5 |
| | 0.975 | 39.0 | 39.2 | 39.2 | 39.3 | 39.3 | 39.4 | 39.4 | 39.4 | 39.4 | 39.4 | 39.4 | 39.5 | 39.5 | 39.5 |
| | 0.990 | 99.0 | 99.2 | 99.2 | 99.3 | 99.3 | 99.4 | 100. | 100. | 100. | 100. | 100. | 100. | 100. | 99.5 |
| | 0.999 | 999. | 999. | | | | | | | | | | | | |
| 3 | 0.900 | 5.46 | 5.39 | 5.34 | 5.31 | 5.28 | 5.27 | 5.25 | 5.23 | 5.22 | 5.20 | 5.18 | 5.17 | 5.15 | 5.13 |
| | 0.950 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.89 | 8.85 | 8.79 | 8.74 | 8.70 | 8.66 | 8.62 | 8.58 | 8.53 |
| | 0.975 | 16.0 | 15.4 | 15.1 | 14.9 | 14.7 | 14.6 | 14.5 | 14.4 | 14.3 | 14.3 | 14.2 | 14.1 | 14.0 | 13.9 |
| | 0.990 | 30.8 | 29.5 | 28.7 | 28.2 | 27.9 | 27.7 | 27.5 | 27.2 | 27.1 | 26.9 | 26.7 | 26.5 | 26.4 | 26.1 |
| | 0.999 | 149. | 141. | 137. | 135. | 133. | 132. | 131. | 129. | 128. | 127. | 126. | 125. | 125. | 123. |
| 4 | 0.900 | 4.32 | 4.19 | 4.11 | 4.05 | 4.01 | 3.98 | 3.95 | 3.92 | 3.90 | 3.87 | 3.84 | 3.82 | 3.79 | 3.76 |
| | 0.950 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.09 | 6.04 | 5.96 | 5.91 | 5.86 | 5.80 | 5.75 | 5.70 | 5.63 |
| | 0.975 | 10.6 | 9.98 | 9.60 | 9.36 | 9.20 | 9.07 | 8.98 | 8.84 | 8.75 | 8.66 | 8.56 | 8.46 | 8.38 | 8.26 |
| | 0.990 | 18.0 | 16.7 | 16.0 | 15.5 | 15.2 | 15.0 | 14.8 | 14.5 | 14.4 | 14.2 | 14.0 | 13.8 | 13.7 | 13.5 |
| | 0.999 | 61.2 | 56.2 | 53.4 | 51.7 | 50.5 | 49.7 | 49.0 | 48.0 | 47.4 | 46.8 | 46.1 | 45.4 | 44.9 | 44.1 |
| 5 | 0.900 | 3.78 | 3.62 | 3.52 | 3.45 | 3.40 | 3.37 | 3.34 | 3.30 | 3.27 | 3.24 | 3.21 | 3.17 | 3.15 | 3.10 |
| | 0.950 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.88 | 4.82 | 4.74 | 4.68 | 4.62 | 4.56 | 4.50 | 4.44 | 4.36 |
| | 0.975 | 8.43 | 7.76 | 7.39 | 7.15 | 6.98 | 6.85 | 6.76 | 6.62 | 6.52 | 6.43 | 6.33 | 6.23 | 6.14 | 6.02 |
| | 0.990 | 13.3 | 12.1 | 11.4 | 11.0 | 10.7 | 10.5 | 10.3 | 10.1 | 9.89 | 9.72 | 9.55 | 9.38 | 9.24 | 9.02 |
| | 0.999 | 37.1 | 33.2 | 31.1 | 29.8 | 28.8 | 28.2 | 27.6 | 26.9 | 26.4 | 25.9 | 25.4 | 24.9 | 24.4 | 23.8 |
| 6 | 0.900 | 3.46 | 3.29 | 3.18 | 3.11 | 3.05 | 3.01 | 2.98 | 2.94 | 2.90 | 2.87 | 2.84 | 2.80 | 2.77 | 2.72 |
| | 0.950 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.21 | 4.15 | 4.06 | 4.00 | 3.94 | 3.87 | 3.81 | 3.75 | 3.67 |
| | 0.975 | 7.26 | 6.60 | 6.23 | 5.99 | 5.82 | 5.70 | 5.60 | 5.46 | 5.37 | 5.27 | 5.17 | 5.07 | 4.98 | 4.85 |
| | 0.990 | 10.9 | 9.78 | 9.15 | 8.75 | 8.47 | 8.26 | 8.10 | 7.87 | 7.72 | 7.56 | 7.40 | 7.23 | 7.09 | 6.88 |
| | 0.999 | 27.0 | 23.7 | 21.9 | 20.8 | 20.0 | 19.5 | 19.0 | 18.4 | 18.0 | 17.6 | 17.1 | 16.7 | 16.3 | 15.7 |
| 7 | 0.900 | 3.26 | 3.07 | 2.96 | 2.88 | 2.83 | 2.78 | 2.75 | 2.70 | 2.67 | 2.63 | 2.59 | 2.56 | 2.52 | 2.47 |
| | 0.950 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.79 | 3.73 | 3.64 | 3.57 | 3.51 | 3.44 | 3.38 | 3.32 | 3.23 |
| | 0.975 | 6.54 | 5.89 | 5.52 | 5.29 | 5.12 | 4.99 | 4.90 | 4.76 | 4.67 | 4.57 | 4.47 | 4.36 | 4.28 | 4.14 |
| | 0.990 | 9.55 | 8.45 | 7.85 | 7.46 | 7.19 | 6.99 | 6.84 | 6.62 | 6.47 | 6.31 | 6.16 | 5.99 | 5.86 | 5.65 |
| | 0.999 | 21.7 | 18.8 | 17.2 | 16.2 | 15.5 | 15.0 | 14.6 | 14.1 | 13.7 | 13.3 | 12.9 | 12.5 | 12.2 | 11.7 |
| 8 | 0.900 | 3.11 | 2.92 | 2.81 | 2.73 | 2.67 | 2.62 | 2.59 | 2.54 | 2.50 | 2.46 | 2.42 | 2.38 | 2.35 | 2.29 |
| | 0.950 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.50 | 3.44 | 3.35 | 3.28 | 3.22 | 3.15 | 3.08 | 3.02 | 2.93 |
| | 0.975 | 6.06 | 5.42 | 5.05 | 4.82 | 4.65 | 4.53 | 4.43 | 4.29 | 4.20 | 4.10 | 4.00 | 3.89 | 3.81 | 3.67 |
| | 0.990 | 8.65 | 7.59 | 7.01 | 6.63 | 6.37 | 6.18 | 6.03 | 5.81 | 5.67 | 5.52 | 5.36 | 5.20 | 5.07 | 4.86 |
| | 0.999 | 18.5 | 15.8 | 14.4 | 13.5 | 12.9 | 12.4 | 12.0 | 11.5 | 11.2 | 10.8 | 10.5 | 10.1 | 9.80 | 9.33 |

# Table of the F-distribution distribution

Entries in table are $F_{1-\alpha}(\nu_1, \nu_2)$: the $\alpha$ tail quantile of the $F_{\nu_1,\nu_2}$ distribution

| $\nu_2 \backslash \nu_l$ | $1-\alpha$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 10 | 12 | 15 | 20 | 30 | 50 | $\infty$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 9 | 0.900 | 3.01 | 2.81 | 2.69 | 2.61 | 2.55 | 2.51 | 2.47 | 2.42 | 2.38 | 2.34 | 2.30 | 2.25 | 2.22 | 2.16 |
| | 0.950 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.29 | 3.23 | 3.14 | 3.07 | 3.01 | 2.94 | 2.86 | 2.80 | 2.71 |
| | 0.975 | 5.71 | 5.08 | 4.72 | 4.48 | 4.32 | 4.20 | 4.10 | 3.96 | 3.87 | 3.77 | 3.67 | 3.56 | 3.47 | 3.33 |
| | 0.990 | 8.02 | 6.99 | 6.42 | 6.06 | 5.80 | 5.61 | 5.47 | 5.26 | 5.11 | 4.96 | 4.81 | 4.65 | 4.52 | 4.31 |
| | 0.999 | 16.4 | 13.9 | 12.6 | 11.7 | 11.1 | 10.7 | 10.4 | 9.89 | 9.57 | 9.24 | 8.90 | 8.55 | 8.26 | 7.81 |
| 10 | 0.900 | 2.92 | 2.73 | 2.61 | 2.52 | 2.46 | 2.41 | 2.38 | 2.32 | 2.28 | 2.24 | 2.20 | 2.16 | 2.12 | 2.06 |
| | 0.950 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.14 | 3.07 | 2.98 | 2.91 | 2.84 | 2.77 | 2.70 | 2.64 | 2.54 |
| | 0.975 | 5.46 | 4.83 | 4.47 | 4.24 | 4.07 | 3.95 | 3.85 | 3.72 | 3.62 | 3.52 | 3.42 | 3.31 | 3.22 | 3.08 |
| | 0.990 | 7.56 | 6.55 | 5.99 | 5.64 | 5.39 | 5.20 | 5.06 | 4.85 | 4.71 | 4.56 | 4.41 | 4.25 | 4.11 | 3.91 |
| | 0.999 | 14.9 | 12.6 | 11.3 | 10.5 | 9.93 | 9.52 | 9.20 | 8.75 | 8.45 | 8.13 | 7.80 | 7.47 | 7.19 | 6.76 |
| 11 | 0.900 | 2.86 | 2.66 | 2.54 | 2.45 | 2.39 | 2.34 | 2.30 | 2.25 | 2.21 | 2.17 | 2.12 | 2.08 | 2.04 | 1.97 |
| | 0.950 | 3.98 | 3.59 | 3.36 | 3.20 | 3.09 | 3.01 | 2.95 | 2.85 | 2.79 | 2.72 | 2.65 | 2.57 | 2.51 | 2.40 |
| | 0.975 | 5.26 | 4.63 | 4.28 | 4.04 | 3.88 | 3.76 | 3.66 | 3.53 | 3.43 | 3.33 | 3.23 | 3.12 | 3.03 | 2.88 |
| | 0.990 | 7.21 | 6.22 | 5.67 | 5.32 | 5.07 | 4.89 | 4.74 | 4.54 | 4.40 | 4.25 | 4.10 | 3.94 | 3.81 | 3.60 |
| | 0.999 | 13.8 | 11.6 | 10.3 | 9.58 | 9.05 | 8.66 | 8.35 | 7.92 | 7.63 | 7.32 | 7.01 | 6.68 | 6.42 | 6.00 |
| 12 | 0.900 | 2.81 | 2.61 | 2.48 | 2.39 | 2.33 | 2.28 | 2.24 | 2.19 | 2.15 | 2.10 | 2.06 | 2.01 | 1.97 | 1.90 |
| | 0.950 | 3.89 | 3.49 | 3.26 | 3.11 | 3.00 | 2.91 | 2.85 | 2.75 | 2.69 | 2.62 | 2.54 | 2.47 | 2.40 | 2.30 |
| | 0.975 | 5.10 | 4.47 | 4.12 | 3.89 | 3.73 | 3.61 | 3.51 | 3.37 | 3.28 | 3.18 | 3.07 | 2.96 | 2.87 | 2.72 |
| | 0.990 | 6.93 | 5.95 | 5.41 | 5.06 | 4.82 | 4.64 | 4.50 | 4.30 | 4.16 | 4.01 | 3.86 | 3.70 | 3.57 | 3.36 |
| | 0.999 | 13.0 | 10.8 | 9.63 | 8.89 | 8.38 | 8.00 | 7.71 | 7.29 | 7.00 | 6.71 | 6.40 | 6.09 | 5.83 | 5.42 |
| 13 | 0.900 | 2.76 | 2.56 | 2.43 | 2.35 | 2.28 | 2.23 | 2.20 | 2.14 | 2.10 | 2.05 | 2.01 | 1.96 | 1.92 | 1.85 |
| | 0.950 | 3.81 | 3.41 | 3.18 | 3.03 | 2.92 | 2.83 | 2.77 | 2.67 | 2.60 | 2.53 | 2.46 | 2.38 | 2.31 | 2.21 |
| | 0.975 | 4.97 | 4.35 | 4.00 | 3.77 | 3.60 | 3.48 | 3.39 | 3.25 | 3.15 | 3.05 | 2.95 | 2.84 | 2.74 | 2.60 |
| | 0.990 | 6.70 | 5.74 | 5.21 | 4.86 | 4.62 | 4.44 | 4.30 | 4.10 | 3.96 | 3.82 | 3.66 | 3.51 | 3.37 | 3.17 |
| | 0.999 | 12.3 | 10.2 | 9.07 | 8.35 | 7.86 | 7.49 | 7.21 | 6.80 | 6.52 | 6.23 | 5.93 | 5.63 | 5.37 | 4.97 |
| 14 | 0.900 | 2.73 | 2.52 | 2.39 | 2.31 | 2.24 | 2.19 | 2.15 | 2.10 | 2.05 | 2.01 | 1.96 | 1.91 | 1.87 | 1.80 |
| | 0.950 | 3.74 | 3.34 | 3.11 | 2.96 | 2.85 | 2.76 | 2.70 | 2.60 | 2.53 | 2.46 | 2.39 | 2.31 | 2.24 | 2.13 |
| | 0.975 | 4.86 | 4.24 | 3.89 | 3.66 | 3.50 | 3.38 | 3.29 | 3.15 | 3.05 | 2.95 | 2.84 | 2.73 | 2.64 | 2.49 |
| | 0.990 | 6.51 | 5.56 | 5.04 | 4.69 | 4.46 | 4.28 | 4.14 | 3.94 | 3.80 | 3.66 | 3.51 | 3.35 | 3.22 | 3.00 |
| | 0.999 | 11.8 | 9.73 | 8.62 | 7.92 | 7.44 | 7.08 | 6.80 | 6.40 | 6.13 | 5.85 | 5.56 | 5.25 | 5.00 | 4.60 |
| 15 | 0.900 | 2.70 | 2.49 | 2.36 | 2.27 | 2.21 | 2.16 | 2.12 | 2.06 | 2.02 | 1.97 | 1.92 | 1.87 | 1.83 | 1.76 |
| | 0.950 | 3.68 | 3.29 | 3.06 | 2.90 | 2.79 | 2.71 | 2.64 | 2.54 | 2.48 | 2.40 | 2.33 | 2.25 | 2.18 | 2.07 |
| | 0.975 | 4.77 | 4.15 | 3.80 | 3.58 | 3.41 | 3.29 | 3.20 | 3.06 | 2.96 | 2.86 | 2.76 | 2.64 | 2.55 | 2.40 |
| | 0.990 | 6.36 | 5.42 | 4.89 | 4.56 | 4.32 | 4.14 | 4.00 | 3.80 | 3.67 | 3.52 | 3.37 | 3.21 | 3.08 | 2.87 |
| | 0.999 | 11.3 | 9.34 | 8.25 | 7.57 | 7.09 | 6.74 | 6.47 | 6.08 | 5.81 | 5.53 | 5.25 | 4.95 | 4.70 | 4.31 |
| 16 | 0.900 | 2.67 | 2.46 | 2.33 | 2.24 | 2.18 | 2.13 | 2.09 | 2.03 | 1.99 | 1.94 | 1.89 | 1.84 | 1.79 | 1.72 |
| | 0.950 | 3.63 | 3.24 | 3.01 | 2.85 | 2.74 | 2.66 | 2.59 | 2.49 | 2.42 | 2.35 | 2.28 | 2.19 | 2.12 | 2.01 |
| | 0.975 | 4.69 | 4.08 | 3.73 | 3.50 | 3.34 | 3.22 | 3.12 | 2.99 | 2.89 | 2.79 | 2.68 | 2.57 | 2.47 | 2.32 |
| | 0.990 | 6.23 | 5.29 | 4.77 | 4.44 | 4.20 | 4.03 | 3.89 | 3.69 | 3.55 | 3.41 | 3.26 | 3.10 | 2.97 | 2.75 |
| | 0.999 | 11.0 | 9.01 | 7.94 | 7.27 | 6.80 | 6.46 | 6.19 | 5.81 | 5.55 | 5.27 | 4.99 | 4.70 | 4.45 | 4.06 |
| 17 | 0.900 | 2.64 | 2.44 | 2.31 | 2.22 | 2.15 | 2.10 | 2.06 | 2.00 | 1.96 | 1.91 | 1.86 | 1.81 | 1.76 | 1.69 |
| | 0.950 | 3.59 | 3.20 | 2.96 | 2.81 | 2.70 | 2.61 | 2.55 | 2.45 | 2.38 | 2.31 | 2.23 | 2.15 | 2.08 | 1.96 |
| | 0.975 | 4.62 | 4.01 | 3.66 | 3.44 | 3.28 | 3.16 | 3.06 | 2.92 | 2.82 | 2.72 | 2.62 | 2.50 | 2.41 | 2.25 |
| | 0.990 | 6.11 | 5.18 | 4.67 | 4.34 | 4.10 | 3.93 | 3.79 | 3.59 | 3.46 | 3.31 | 3.16 | 3.00 | 2.87 | 2.65 |
| | 0.999 | 10.7 | 8.73 | 7.68 | 7.02 | 6.56 | 6.22 | 5.96 | 5.58 | 5.32 | 5.05 | 4.77 | 4.48 | 4.24 | 3.85 |