University of British Columbia
Final Examination
STAT 305 Introduction to Statistical Inference 2015–16 Term 2
Instructor: William J. Welch

Student Family Name:	______________________ (Please PRINT)
Student Given Names:	______________________ (Please PRINT)
Student ID Number:	______________________
Signature:	______________________

Date of Exam:	*April 14, 2016*
Time Period:	*7:00–9:30 pm*
Number of Exam Pages:	*13, including this cover sheet (please check for completeness)*
Additional Materials Allowed:	*Calculator; formula sheet* ($8\frac{1}{2} \times 11$, *2-sided)*

Question	Marks	Score
1	15	
2	13	
3	14	
4	14	
Total	56	

Student Conduct During Examinations

1. Each examination candidate must be prepared to produce, upon request of the invigilator or examiner, his or her UBCcard for identification.

2. Examination candidates are not permitted to ask questions of the examiners or invigilators, except in cases of supposed errors or ambiguities in examination questions, illegible material, or the like.

3. No examination candidate shall be permitted to enter the examination room after the expiration of one-half hour from the scheduled starting time, or to leave during the first half hour of the examination. Should the examination run forty-five (45) minutes or less, no examination candidate shall be permitted to enter the examination room once the examination has begun.

4. Examination candidates must conduct themselves honestly and in accordance with established rules for a given examination, which will be articulated by the examiner or invigilator prior to the examination commencing. Should dishonest behaviour be observed by the examiner(s) or invigilator(s), pleas of accident or forgetfulness shall not be received.

5. Examination candidates suspected of any of the following, or any other similar practices, may be immediately dismissed from the examination by the examiner/invigilator, and may be subject to disciplinary action:

   (a) speaking or communicating with other candidates, unless otherwise authorized;

   (b) purposely exposing written papers to the view of other examination candidates or imaging devices;

   (c) purposely viewing the written papers of other examination candidates;

   (d) using or having visible at the place of writing any books, papers or other memory aid devices other than those authorized by the examiner(s); and,

   (e) using or operating electronic devices including but not limited to telephones, calculators, computers, or similar devices other than those authorized by the examiner(s)—(electronic devices other than those authorized by the examiner(s) must be completely powered down if present at the place of writing).

6. Examination candidates must not destroy or damage any examination material, must hand in all examination papers, and must not take any examination material from the examination room without permission of the examiner or invigilator.

7. Notwithstanding the above, for any mode of examination that does not fall into the traditional, paper-based method, examination candidates shall adhere to any special rules for conduct as established and articulated by the examiner.

8. Examination candidates must follow any additional examination rules or directions communicated by the examiner(s) or invigilator(s).

Tables 1 and 2 at the end of the booklet contain some common discrete and continuous distributions, along with their properties. They are the same as Tables 1.5 and 1.6 in the Course Notes.

For full marks in questions asking you to show or derive a property, you must be clear about any result you are using, INCLUDING ANY CONDITIONS FOR THE RESULT TO HOLD, and how the result is applied.

1. Sample data, y, on the number of Bernoulli "trials" required for n "successes" is assumed to be a value of $Y \sim \mathsf{NegBin}\,(n, \pi)$. The objective here is to estimate π via Bayes' Rule. The random variable Π is introduced to represent uncertainty about π. The prior distribution for Π is taken to be $\mathsf{Beta}\,(a, b)$.

   (a) [3 marks] Show that the Bayesian posterior distribution of Π given data y is a beta distribution with its parameters taking values $a + n$ and $b + y - n$, respectively.

   (b) [2 marks] Is the beta distribution a conjugate prior here? Why or why not?

   (c) Patients with high cholesterol entering a medical study are given an experimental drug. Each patient has his/her cholesterol status assessed as still "elevated" or not after one week of treatment. Recruitment of patients stops when $n = 64$ patients are assessed to have elevated cholesterol; this requires $y = 296$ patients to enter the study. It is assumed that y is the value of $Y \sim \mathsf{NegBin}\,(n, \pi)$, where π is the probability of elevated cholesterol. The prior for Π is $\mathsf{Unif}\,(0, 1)$.

      i. [1 mark] What is the posterior distribution of Π?

      ii. [1 mark] Numerically, what is the posterior expectation of Π?

iii. [2 marks] Suppose the posterior expectation is taken as an estimate of the true value of π. Does this estimate make intuitive sense? Why or why not?

iv. It can be shown that a 95% credibility interval of π values is $[0.173, 0.267]$. Maximum likelihood (ML), based on the same negative-binomial probability model and the same data, gives a 95% confidence interval $[0.169, 0.263]$ for π. For each of the following criteria, say whether the Bayesian or ML method has an advantage relative to the other method and give a brief reason.

A. [1.5 marks] Consideration of sample data that might have occurred but did not.

B. [1.5 marks] Mathematical exactness or approximation in the derivation of the interval.

C. [1.5 marks] Objectivity of the analysis.

D. [1.5 marks] An intuitive interpretation of the interval.

2. Judd et al. ("Acute Operative Stabilization Versus Nonoperative Management of Clavicle Fractures", *American Journal of Orthopedics*, 2009, pp. 341–345) compared two treatments for a fractured clavicle (broken collarbone): operative and nonoperative. They divided 57 volunteer patients into operative and nonoperative groups by use of sealed envelopes in a random order. This led to 29 operative and 28 nonoperative patients.

   Various scores measuring the success of the treatments were presented. Here we consider the SANE measure (y) at 1 year after injury. Larger y values correspond to a greater degree of function of the injured shoulder.

   Sample means and standard deviations of the observed y scores for the two groups are given in the following table.

		Sample	
Group	n	mean	sd
Operative	29	93.5	4.2
Nonoperative	28	97.0	3.6

   (a) [1 mark] Why were the 57 participating patients randomized to the two treatments?

   (b) [1 mark] The study was not paired. If it had been paired, speculate briefly on how the study might have been conducted.

   (c) The y scores in the operative sample are assumed to be independent draws from a $\mathsf{N}(\mu_A, \sigma^2)$ distribution. Similarly, the scores in the nonoperative sample are assumed to be independent realizations from a $\mathsf{N}(\mu_B, \sigma^2)$ distribution. Note that the variance, σ^2, is taken to be the same for the two distributions, but it is unknown. Furthermore, the two samples are statistically independent of each other.

      i. [1 mark] Compute a numerical estimate of $\delta = \mu_A - \mu_B$.

      ii. [2 marks] Compute a numerical estimate of σ^2.

iii. [2 marks] Compute a standard error for the estimator of δ.

iv. [2 marks] Compute a 98% two-sided confidence interval for δ. You might find the following `R` output useful:

```
> qt(0.01, df = 56)
[1] -2.394801
> qt(0.02, df=56)
[1] -2.102696
> qt(0.01, df=55)
[1] -2.396081
> qt(0.02, df=55)
[1] -2.103607
```

v. [2 marks] Judd et al. carried out a hypothesis test of $H_0 : \delta = 0$ versus $H_a : \delta \neq 0$. Suppose they used a significance level of 2%. Without doing any further calculations, what is the outcome of the test? Explain.

vi. [2 marks] Suppose a difference in means of $\delta = +5$ is medically important. For simplicity, further assume $\sigma = 4$ is known, so that a test based on the normal distribution can be used. Write down an expression for the power of the hypothesis test of $H_0 : \delta = 0$ versus $H_a : \delta \neq 0$ at significance level 2% when $\delta = 5$ and $\sigma = 4$.

3. Lappe et al. (*American Journal of Clinical Nutrition*, 2007, pp. 1586–1591) studied the effect of two dietary supplements on cancer rates. Participants were allocated at random to two samples; subjects in the first sample received a supplement of calcium only, while those in the second received a supplement of vitamin D plus calcium. (Another group receiving a placebo with neither supplement is ignored in this question.) Participants were monitored for various types of cancers for four years, to give the data in the following table.

	Number of subjects		
	Ca only	Vit. D + Ca	Total
Breast cancer	6	5	11
Other cancers	11	8	19
No cancer	428	433	861
Total	445	446	891

For example, of the 445 subjects in the calcium-only sample, 6 developed breast cancer, 11 had cancers of all other types, and 428 developed no cancer.

(a) For the calcium-only treatment, assume that the three observed frequencies are randomly drawn from a multinomial distribution over three categories (breast cancer, other cancers, and no cancer) with category probabilities π_{11}, π_{21}, and π_{31}, respectively. Similarly, assume the observed frequencies for vitamin D plus calcium are randomly drawn from a multinomial distribution over the same categories but with probabilities π_{12}, π_{22}, and π_{32}, respectively.

i. [1 mark] Write down a mathematical definition of the null hypothesis, H_0, that the probability of breast cancer is the same across the two treatments, similarly the probability of other cancers is the same, and the probability of no cancer is the same.

ii. [1 mark] What is the alternative hypothesis, H_a?

iii. [1 mark] How many free parameters are there among π_{11}, π_{21}, π_{31}, π_{12}, π_{22}, and π_{32} under H_0?

iv. [1 mark] How many free parameters are there among π_{11}, π_{21}, π_{31}, π_{12}, π_{22}, and π_{32} if H_0 is not imposed?

v. [2 marks] Suppose H_0 is tested by computing Pearson's statistic. From what approximate distribution is the value of the statistic drawn if H_0 is true?

(b) Now assume that the six observed frequencies y_{ij} for $i = 1, 2, 3$ and $j = 1, 2$ are frequencies for a random sample of size 891 from one multinomial distribution with six category probabilities π_{ij}. The six categories are defined by a row variable with index i for breast cancer, other cancers, and no cancer, respectively, and a column variable with index j for the Ca and vitamin D + Ca treatments, respectively.

i. [1 mark] A random draw from this multinomial distribution will place a subject in one of the six categories. Write down a mathematical definition of the null hypothesis, H_0, that the row in which she falls is statistically independent of the column in which she falls.

ii. [1 mark] What is the alternative hypothesis, H_a?

iii. [2 marks] How many free parameters are there among π_{11}, π_{21}, π_{31}, π_{12}, π_{22}, and π_{32} under H_0? Explain.

iv. [1 mark] How many free parameters are there among π_{11}, π_{21}, π_{31}, π_{12}, π_{22}, and π_{32} if H_0 is not imposed?

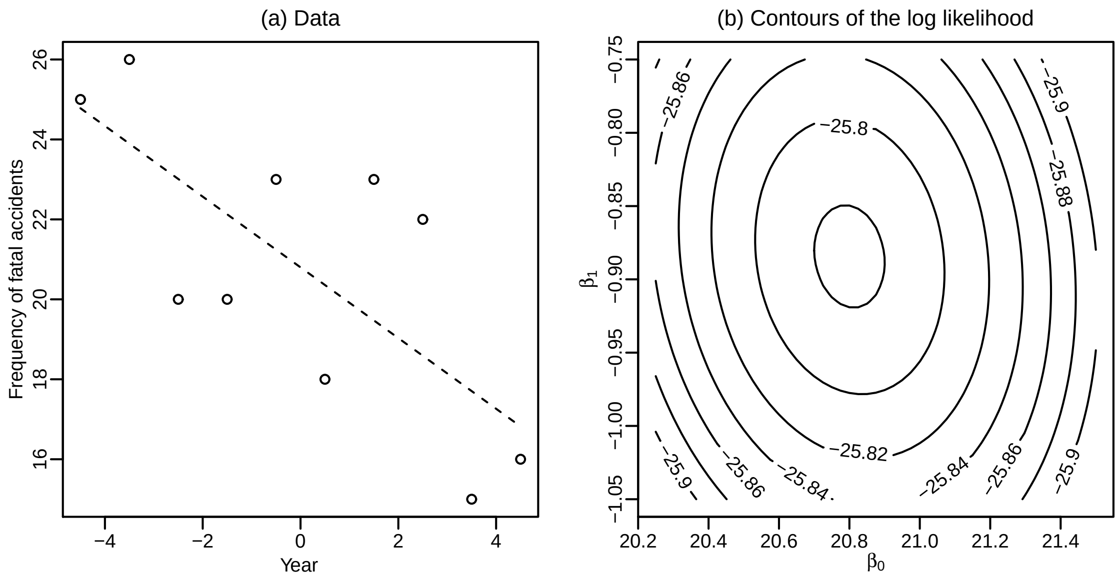
v. [1 mark] Suppose H_0 is tested by computing Pearson's statistic. From what approximate distribution is the value of the statistic drawn if H_0 is true?

(c) [2 marks] Suppose the same significance level is used for the hypothesis tests outlined in questions 3a and 3b. Will the two tests give the same result in terms of rejecting H_0 or not? Explain why or why not.

4. The International Air Transportation Association (IATA) *Safety Report 2013* gives data on frequencies of fatal accidents (y) for the 10 years 2004–2013, as in the following table.

Year	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Accidents	25	26	20	20	23	18	23	22	15	16

The data are shown in the left panel below. Year has been centred by subtracting the mean year to make numerical calculations easier; the conclusions are not affected. We use the notation x for the adjusted year.



This question treats the 10 y_i values as realizations of independent random variables $Y_1, \ldots, Y_{10}$, where $Y_i \sim \textsf{Pois}\,(\mu_i)$. The mean number of accidents follows a trend with year:

$$\mu_i = \beta_0 + \beta_1 x_i \quad (-\infty < \beta_0, \beta_1 < \infty).$$

Note that μ_i (and not $\ln(\mu_i)$) follows a linear trend with year. The fitted line drawn in the left panel above is from maximum likelihood (ML) estimation, the topic of this question.

(a) [1 mark] State one possible disadvantage of modelling the trend as $\mu_i = \beta_0 + \beta_1 x_i$ versus $\ln(\mu_i) = \beta_0 + \beta_1 x_i$. (Nonetheless, use $\mu_i = \beta_0 + \beta_1 x_i$ for remaining parts of this question.)

(b) [2 marks] The right panel above shows contours where the joint log likelihood function takes various constant values. (The contour in the centre that `R` chose not to label has value -25.785.) Approximately, what are the numerical maximum likelihood estimates of β_0 and β_1?

(c) [2 marks] Obtain an expression for the joint log-likelihood as a function of β_0 and β_1.

(d) [3 marks] It can be shown that the Fisher information for estimating both β_0 and β_1 is

$$\begin{pmatrix} \sum_{i=1}^{10}(1/\mu_i) & \sum_{i=1}^{10}(x_i/\mu_i) \\ \sum_{i=1}^{10}(x_i/\mu_i) & ?? \end{pmatrix}.$$

Derive the formula for the missing element in the matrix.

(e) [2 marks] State briefly how you would use the Fisher information matrix to find an approximate formula for $\mathrm{Var}(\tilde{\beta}_1)$.

(f) [2 marks] State briefly how you would use the approximate formula for $\mathrm{Var}(\tilde{\beta}_1)$ to find a standard error for $\tilde{\beta}_1$.

(g) [2 marks] If `R` is used as a matrix calculator, it can be shown that the standard error of $\tilde{\beta}_1$ is 0.499. `R` would also give the following output.

```
> qnorm(0.95)
[1] 1.644854
> qnorm(0.975)
[1] 1.959964
```

Compute an approximate, two-sided 95% confidence interval for β_1.

Distribution and notation	PMF, $f_Y(y)$	$\mathrm{E}(Y)$	$\mathrm{Var}(Y)$	MGF, $M_Y(t)$
Bernoulli	$f_Y(0) = 1 - \pi,\ f_Y(1) = \pi$	π	$\pi(1-\pi)$	$1 - \pi + \pi e^t$
Bern (π)	$(y = 0, 1;\ 0 < \pi < 1)$			$(-\infty < t < \infty)$
Binomial	$\binom{n}{y}\pi^y(1-\pi)^{n-y}$	$n\pi$	$n\pi(1-\pi)$	$(1 - \pi + \pi e^t)^n$
Bin (n, π)	$(y = 0, 1, \ldots, n;$ $n = 1, 2, \ldots;\ 0 < \pi < 1)$			$(-\infty < t < \infty)$
Geometric (two versions)				
Geom0 (π)	$(1-\pi)^y\pi$ $(y = 0, 1, \ldots, \infty;$ $0 < \pi < 1)$	$\frac{1-\pi}{\pi}$	$\frac{1-\pi}{\pi^2}$	$\frac{\pi}{1-(1-\pi)e^t}$ $(-\infty < t <$ $-\ln(1-\pi))$
Geom1 (π)	$(1-\pi)^{y-1}\pi$ $(y = 1, 2, \ldots, \infty;$ $0 < \pi < 1)$	$\frac{1}{\pi}$	$\frac{1-\pi}{\pi^2}$	$\frac{e^t\pi}{1-(1-\pi)e^t}$ $(-\infty < t <$ $-\ln(1-\pi))$
Negative binomial NegBin (n, π)	$\binom{y-1}{n-1}(1-\pi)^{y-n}\pi^n$ $(y = n, n+1, \ldots, \infty;$ $n = 1, 2, \ldots, \infty;$ $0 < \pi < 1)$	$\frac{n}{\pi}$	$\frac{n(1-\pi)}{\pi^2}$	$\left(\frac{e^t\pi}{1-(1-\pi)e^t}\right)^n$ $(-\infty < t <$ $-\ln(1-\pi))$
Poisson	$\frac{e^{-\mu}\mu^y}{y!}$	μ	μ	$e^{\mu(e^t-1)}$
Pois (μ)	$(y = 0, 1, \ldots, \infty;\ \mu > 0)$			$(-\infty < t < \infty)$

Table 1: Some commonly used discrete distributions, along with their expectations, variances, and moment generating functions (MGFs).

Distribution and notation	PDF, $f_Y(y)$	E(Y)	Var(Y)	MGF, $M_Y(t)$
Beta $\mathsf{Beta}\,(a,b)$	$\frac{1}{B(a,b)}y^{a-1}(1-y)^{b-1}$ $(0<y<1;$ $a>0;\ b>0)$	$\frac{a}{a+b}$	$\frac{ab}{(a+b)^2(a+b+1)}$	Not useful
Chi-squared χ^2_d	$\frac{1}{2^{d/2}\Gamma(d/2)}y^{d/2-1}e^{-y/2}$ $(0<y<\infty;$ $d=1,2,\ldots)$	d	$2d$	$\frac{1}{(1-2t)^{d/2}}$ $(-\infty<t<\frac{1}{2})$
Exponential $\mathsf{Expon}\,(\lambda)$	$\lambda e^{-\lambda y}$ $(0<y<\infty;\ \lambda>0)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-t}$ $(-\infty<t<\lambda)$
Fisher's F F_{d_1,d_2}	$\frac{(d_1/d_2)^{d_1/2}y^{d_1/2-1}}{B\left(\frac{d_1}{2},\frac{d_2}{2}\right)\left(1+\frac{d_1}{d_2}y\right)^{\frac{d_1+d_2}{2}}}$ $(0<y<\infty;$ $d_1,d_2=1,2,\ldots)$	$\frac{d_2}{d_2-2}$ $(d_2>2)$	$\frac{2d_2^2(d_1+d_2-2)}{d_1(d_2-2)^2(d_2-4)}$ $(d_2>4)$	Does not exist
Gamma $\mathsf{Gamma}\,(\nu,\lambda)$	$\frac{1}{\Gamma(\nu)}\lambda(\lambda y)^{\nu-1}e^{-\lambda y}$ $(0<y<\infty;$ $\nu>0;\ \lambda>0)$	$\frac{\nu}{\lambda}$	$\frac{\nu}{\lambda^2}$	$\left(\frac{\lambda}{\lambda-t}\right)^{\nu}$ $(-\infty<t<\lambda)$
Normal $\mathsf{N}\,(\mu,\sigma^2)$	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}(y-\mu)^2}$ $(-\infty<y<\infty;$ $-\infty<\mu<\infty;\ \sigma^2>0)$	μ	σ^2	$e^{\mu t+\frac{1}{2}\sigma^2t^2}$ $(-\infty<t<\infty)$
Student's t t_d	$\frac{1}{B\left(\frac{1}{2},\frac{d}{2}\right)\sqrt{d}\left(1+\frac{y^2}{d}\right)^{\frac{d+1}{2}}}$ $(-\infty<y<\infty;$ $d=1,2,\ldots)$	0 $(d>1)$	$\frac{d}{d-2}$ $(d>2)$	Does not exist
Uniform (rectangular) $\mathsf{Unif}\,(a,b)$	$\frac{1}{b-a}$ $(a<y<b;\ a<b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{bt}-e^{at}}{(b-a)t}$ $(-\infty<t<\infty)$

Table 2: Some commonly used continuous distributions, along with their expectations, variances, and moment generating functions (MGFs).