out of 10

Name Jerry

Student Number

6

STA 442/2101 f2014 Quiz 1

4

1. (5 points) A random sample of size n was drawn from a distribution with density $f(y) = \theta e^{-\theta y}$ for $y > 0$, where the parameter $\theta > 0$.

   (a) Find the maximum likelihood estimator of θ. Show your work. Don't bother wth a second derivative test. Your answer is a symbolic expression. **Circle your final answer.**

$$\ell(\theta) = \log \prod_{i=1}^{n} \theta e^{-\theta y_i} = \log\left(\theta^n e^{-\theta \sum y_i}\right)$$

$$= n \log \theta - \theta \sum_{i=1}^{n} y_i$$

$$\frac{d\ell}{d\theta} = \frac{n}{\theta} - \sum y_i \overset{set}{=} 0 \Rightarrow \frac{n}{\theta} = \sum y_i$$

$$\Rightarrow \boxed{\theta = \frac{n}{\sum_{i=1}^{n} y_i} = \hat{\theta}}$$

1

   (b) The data values are 4.1, 9.3, 2.2, 4.4. Give the maximum likelihood estimate in numeric form. Your answer is a number. **Circle it.**

$$\hat{\theta} = \frac{4}{20} = 1/5$$

4

2. (5 points) Let $\mathbf{X}$ be a real $n \times p$ matrix, and let $\mathbf{a}$ be a real $p \times 1$ column vector. Show that $\mathbf{a}'(\mathbf{X}'\mathbf{X})\mathbf{a} \geq 0$ (that is, $\mathbf{X}'\mathbf{X}$ is non-negative definite). You have a lot more room than you need.

$$a'(x'x)a = (a'x')xa$$

$$= \underbrace{(xa)'}_{\in \mathbb{R}^{n\times 1}} \underbrace{xa}_{\in \mathbb{R}^{n\times 1}} = \sum_{j=1}^{n} z_j^2 \geq 0$$

Name Jerry

Student Number

STA 442/2101 f2014 Quiz 2

1. (5 points) Independently for $i = 1, \ldots, n$,

$$
\begin{aligned}
X_i &= \delta + Z_i + \epsilon_{i1} \\
Y_i &= Z_i + \epsilon_{i2}, \text{ where}
\end{aligned}
$$

- δ is a constant.
- $E(Z_i) = \mu_z$, $Var(Z_i) = \sigma_z^2 > 0$
- $E(\epsilon_{i1}) = E(\epsilon_{i2}) = 0$,
- $Var(\epsilon_{i1}) = \sigma_1^2 > 0$, $Var(\epsilon_{i2}) = \sigma_2^2 > 0$
- Z_i and ϵ_{ij} are all independent.

Prove $Cov(X_i, Y_i) > 0$.

$$Cov(X_i, Y_i) = E(X_i Y_i) - E(X_i)E(Y_i)$$
$$= E(X_i Y_i) - (\delta + \mu_z)\mu_z \text{, and}$$

$$E(X_i Y_i) = E(\delta + Z_i + \varepsilon_{i1})(Z_i + \varepsilon_{i2})$$
$$= E(\delta Z_i + \delta \varepsilon_{i2} + Z_i^2 + Z_i \varepsilon_{i2} + \varepsilon_{i1} Z_i + \varepsilon_{i1}\varepsilon_{i2})$$
$$= \delta E(Z_i) + \delta E(\varepsilon_{i2}) + E(Z_i^2) + E(Z_i)E(\varepsilon_{i2}) + E(\varepsilon_{i1})E(Z_1) + E(\varepsilon_{i1})E(\varepsilon_{i2})$$
$$= \delta \cdot \mu_z + \delta \cdot 0 + (\sigma_z^2 + \mu_z^2) + \mu_z \cdot 0 + 0 \cdot \mu_z + 0 \cdot 0$$
$$= \delta \mu_z + \sigma_z^2 + \mu_z^2$$

$$Cov(X_i, Y_i) = \cancel{\delta \mu_z} + \sigma_z^2 + \cancel{\mu_z^2} \cancel{- \delta \mu_z} \cancel{- \mu_z^2}$$
$$= \sigma_z^2 > 0$$

2. **Please slow down and read this question carefully**. Let $Y_1, \ldots, Y_n$ be a random sample from a distribution with expected value μ and variance σ^2. To test $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$, we will use the test statistic $Z_1 = \frac{\sqrt{n}(\overline{Y}-\mu_0)}{S}$, where S is the sample standard deviation. The null hypothesis will be rejected if $Z_1 > z_\alpha$, where z_α is the point that cuts off the top α of the standard normal distribution. Notice this is a one-sided test.

Also notice that by the Central Limit Theorem, $Z_2 = \frac{\sqrt{n}(\overline{Y}-\mu)}{\sigma}$ is approximately standard normal regardless of whether H_0 is true or not.

(a) (4 points) **Give an expression for the approximate *power* of the test** — that is, the probability of rejecting H_0 when H_0 is false. Because this is a large sample test, assume n is big enough so you can replace S with σ any time you wish, and probabilities remain roughly the same. Write your answer in terms of Φ, the cumulative distribution function of a standard normal. The expression also involves μ, σ^2, z_α and n. Show your work. Simplify. **Circle your final answer.**

$$\text{Power} = Pr\{Z_1 > z_\alpha\} = Pr\left\{\frac{\sqrt{n}(\overline{Y}-\mu_0)}{S} > z_\alpha\right\}$$

$$\approx Pr\left\{\frac{\sqrt{n}(\overline{Y}-\mu_0)}{\sigma} > z_\alpha\right\} = Pr\left\{\overline{Y}-\mu_0 > z_\alpha \frac{\sigma}{\sqrt{n}}\right\}$$

$$= Pr\left\{\overline{Y} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}\right\} = Pr\left\{\overline{Y}-\mu > \mu_0 - \mu + z_\alpha \frac{\sigma}{\sqrt{n}}\right\}$$

$$= Pr\left\{\frac{\sqrt{n}(\overline{Y}-\mu)}{\sigma} > \frac{\sqrt{n}(\mu_0-\mu)}{\sigma} + \frac{\cancel{\sqrt{n}}}{\cancel{\sigma}} z_\alpha \frac{\cancel{\sigma}}{\cancel{\sqrt{n}}}\right\}$$

$$= Pr\left\{Z_2 > \frac{\sqrt{n}(\mu_0-\mu)}{\sigma} + z_\alpha\right\}$$

$$= 1 - \Phi\left(\frac{\sqrt{n}(\mu_0-\mu)}{\sigma} + z_\alpha\right)$$

(b) (1 point) When $\mu > \mu_0$, what happens to the power as the sample size $n \to \infty$? Use your answer to Question 2a. Show your work. Hint: The function Φ is continuous, so the limit of the function is the function of the limit.

$$\lim_{n\to\infty}\left(1 - \Phi\left(\frac{\sqrt{n}(\mu_0-\mu)}{\sigma} + z_\alpha\right)\right) = 1 - \Phi\left(\lim_{n\to\infty}\frac{\sqrt{n}(\mu_0-\mu)}{\sigma} + z_\alpha\right)$$

$$= 1 - \Phi(-\infty) = 1 - 0 = 1$$

Kind of rough but okay

Name Jerry

Student Number

STA 442/2101 f2014 Quiz 3

1. (3 points) In this simple regression through the origin, values of the explanatory variable are random, not fixed constants. Independently for $i = 1, \ldots, n$, let $Y_i = \beta X_i + \epsilon_i$, where $E(X_i) = E(\epsilon_i) = 0$, $Var(X_i) = \sigma_x^2$, $Var(\epsilon_i) = \sigma^2$, and ϵ_i is independent of X_i. Let $\widehat{\beta}_n = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2}$. Is $\widehat{\beta}_n$ a consistent estimator of β? Answer Yes or No and prove your answer.

$$E(X_i Y_i) = E\{X_i(\beta X_i + \varepsilon_i)\}$$

$$= \beta E(X_i^2) + E(X_i \varepsilon_i) = \beta \sigma_x^2 + E(X_i)E(\varepsilon_i)$$

$$= \beta \sigma_x^2 + 0$$

And $E(X_i^2) = \sigma_x^2 + 0^2 = \sigma_x^2$,

So $\frac{1}{n}\sum_{i=1}^n X_i Y_i \xrightarrow{as} \beta\sigma_x^2$ by LLN and

$\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{as} \sigma_x^2$ by LLN, and

by continuous mapping,

$$\widehat{\beta}_n = \frac{\frac{1}{n}\sum_{i=1}^n X_i Y_i}{\frac{1}{n}\sum_{i=1}^n X_i^2} \xrightarrow{as} \frac{\beta \cancel{\sigma_x^2}}{\cancel{\sigma_x^2}} = \beta$$

Consistent, YES (strongly consistent)

Name Jerry

Student Number ____________

STA 442/2101 f2014 Quiz 6

1. For a simple linear regression with an intercept and one explanatory variable, you obtain the least squares estimates by minimizing $Q(\beta_0, \beta_1) = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_i)^2$.

   (a) (2 points) Differentiate Q with respect to β_0 and set the derivative to zero and simplify a bit, obtaining the first *normal equation.*

$$\frac{\partial Q}{\partial \beta_0} = \sum_{i=1}^{n} 2(Y_i - \beta_0 - \beta_1 x_i)(-1) \stackrel{set}{=} 0$$

$$\Rightarrow \sum_{i=1}^{n} Y_i = n\beta_0 + \beta_1 \sum_{i=1}^{n} x_i$$

   (b) (3 points) Noting that the quantities $\widehat{\beta}_0$ and $\widehat{\beta}_1$ must satisfy the first normal equation, show that the least squares line passes through the point $(\overline{x}, \overline{Y})$. *Start by giving the equation of the least squares line* in terms of $\widehat{\beta}_0$ and $\widehat{\beta}_1$.

Equation of the least squares line is $y = \hat{\beta}_0 + \hat{\beta}_1 x$.

From the first normal equation,

$$\sum_{i=1}^{n} Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i \Rightarrow \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{n}{n}\hat{\beta}_0 + \hat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n}$$

$$\Rightarrow \overline{Y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{x}$$ and the line goes through $(\underline{x}, \overline{Y})$

2. (2 points) The density of the t distribution is symmetric around zero, so that if $T \sim t(\nu)$, then also $-T \sim t(\nu)$. This fact makes it possible for a normal person to just write down confidence intervals and prediction intervals from the formula sheet without showing any work. Based on a regression with n cases, give a $(1-\alpha)100\%$ prediction interval for Y_{n+1}. Use $t_{\alpha/2}$ for the point satisfying $P\{T > t_{\alpha/2}\} = \alpha/2$. *You don't have to show any work.* Give it in the form $1-\alpha$ equals the probability of something.

$$1-\alpha = Pr\Big\{ x_{n+1}^T \hat{\beta} - t_{\alpha/2}\sqrt{MSE\,(1 + x_{n+1}^T (X^TX)^{-1} x_{n+1}^T)}$$
$$< Y_{n+1} <$$
$$x_{n+1}^T \hat{\beta} + t_{\alpha/2}\sqrt{MSE\,(1 + x_{n+1}^T (X^TX)^{-1} x_{n+1})}\Big\}$$

3. (3 points) Of course the explanatory variable values are often random, and *not* fixed constants. Suppose the explanatory variables are random. Is the prediction interval still valid? **Answer Yes or No** and show your work. For convenience of notation, you may pretend that the joint distribution of the **X** matrix and x_{n+1} is continuous.

To avoid a lot of writing, let $A(W)$ denote the lower limit of your prediction interval, and let $B(W)$ denote the upper limit.

$$Pr\{A(W) < Y_{n+1} < B(W)\}$$

$$= \int \cdots \int P\{A(W) < Y_{n+1} < B(W) \mid W = w\}\, f(w)\, dw$$

$$= \int \cdots \int (1-\alpha)\, f(w)\, dw = (1-\alpha) \int \cdots \int f(w)\, dw$$

$$= (1-\alpha)\cdot 1 = 1-\alpha$$

<u>YES</u>

Name Jerry

Student Number

STA 442/2101 f2014 Quiz 7

1. (6 points) In a study comparing the effectiveness of different weight loss diets, volunteers were randomly assigned to one of two diets (A or B) or put on a waiting list and advised to lose weight on their own. Participants were weighed before and after 6 months of participation in the program (or 6 months of being on the waiting list). The response variable is weight *loss*. The explanatory variables are age (a covariate) and treatment group.

   (a) Write the regression equation. Your model should have *no intercept*, and *parallel regression lines*. Please use x for age. You don't have to say how your dummy variables are defined. You'll do that in the next part.

   $$Y_i = \beta_1 d_{i1} + \beta_2 d_{i2} + \beta_3 d_{i3} + \beta_4 x_i + \epsilon_i$$

   Okay to give $E(Y|x,d)$

   (b) Make a table with three rows, showing how you would set up indicator dummy variables for treatment group. Give $E(Y|x)$ in the last column.

	d_1	d_2	d_3	$E(Y|x)$
A	1	0	0	$\beta_1 + \beta_4 x$
B	0	1	0	$\beta_2 + \beta_4 x$
Control	0	0	1	$\beta_3 + \beta_4 x$

   (c) In terms of β values, what null hypothesis would you test to find out whether, allowing for age, the three diets (including Wait List) differ in their effectiveness?

   $H_0: \beta_1 = \beta_2 = \beta_3$

   (d) In terms of β values, what null hypothesis would you test to find out whether, allowing for age, diets A and B differ in their effectiveness?

   $H_0: \beta_1 = \beta_2$

   (e) In terms of β values, what null hypothesis would you test to find out whether the Wait list "diet" is of any value at all in helping 25-year-old participants to lose weight? Remember, Y is weight loss, which could be zero or even negative.

   $H_0: \beta_3 + 25\beta_4 = 0$

   (f) Is it safe to assume that age is independent of treatment group? Answer Yes or No and briefly explain.

   Yes, because of random assignment

Name Jenny

Student Number

STA 442/2101 f2016 Quiz 1

1. (4 points) The $n \times n$ real matrix $\mathbf{A}$ is said to be non-negative definite if $\mathbf{v}^\top \mathbf{A} \mathbf{v} \geq 0$ for every real $n \times 1$ vector $\mathbf{v}$. Suppose that $\mathbf{A}$ is a *symmetric* matrix that is also *idempotent*, meaning $\mathbf{AA} = \mathbf{A}$. Prove that $\mathbf{A}$ is non-negative definite.

$$v^T A v = v^T A A v = v^T A^T A v$$

$$= \underset{1 \times n}{(Av)^T} \underset{n \times 1}{Av} = z^T T = \sum_{i=1}^{n} z_i^2 \geq 0$$

2. (2 points) In a clinical trial comparing an new treatment for arthritis to an existing treatment, we test $H_0 : \mu_1 = \mu_2$ and obtain $p > 0.05$. **True or False**: We conclude that the two treatments are not equally effective. Just write the word "True" or the word "False" in the space below.

False

3. (4 points) Homework Problem 7 was about power and sample size for a one-sided test of the variance. In the last part of the question, you were asked to calculate the smallest n that yields a power of at least 0.8 when $\sigma^2 = 5$. **Write the required sample size (a number from your printout) in the space below. On your printout, circle the number and write "Question 3" beside it.**

$n = 245$

Please **attach your R printout to the quiz.** Make sure your name and student number are on your printout. *If you do not have your R printout, do not answer Question 3.*

Name Jerry

Student Number

STA 442/2101 f2016 Quiz One

1. (4 points) The $n \times n$ real matrix $\mathbf{A}$ is said to be *positive definite* if $\mathbf{v}^\top \mathbf{A} \mathbf{v} > 0$ for every real $n \times 1$ vector $\mathbf{v} \neq \mathbf{0}$. Suppose $\mathbf{A}$ is a 3×3 matrix. Prove that element $(1, 1)$ — that is, the upper left element — is positive.

$$v^T A v = (1\ 0\ 0) A \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = a_{11} > 0$$

2. (2 points) In a clinical trial comparing a new treatment for arthritis to an existing treatment, we test $H_0 : \mu_1 = \mu_2$ and obtain $p < 0.05$. **True or False**: We conclude that the two treatments are not equally effective. Just write the word "True" or the word "False" in the space below.

True

3. (4 points) Homework Problem 7 was about power and sample size for a one-sided test of the variance. In the last part of the question, you were asked to calculate the smallest n that yields a power of at least 0.8 when $\sigma^2 = 5$. **Write the required sample size (a number from your printout) in the space below. On your printout, circle the number and write "Question 3" beside it.**

$n = 245$

Please **attach your R printout to the quiz**. Make sure your name and student number are on your printout. *If you do not have your R printout, do not answer Question 3.*

Name Jerry

Student Number

STA 442/2101 f2016 Quiz 2

1. (5 points) A web page has advertisements on a side panel. Visitors to the site are randomly assigned to see either ad A or ad B. Whether they click on the ad is recorded, Yes or No. If a person visits the site more than once during the brief study period, only the data from the first visit are recorded.

   (a) State a reasonable model for these data. *You do not have to justify your model* (you may make a few remarks if you wish), but the data have certain features that must be reflected in your model for full marks.

$$\left.\begin{array}{l} x_1, \ldots x_{n_1} \stackrel{iid}{\sim} B(1, \theta_1) \\ y_1, \ldots y_{n_2} \stackrel{iid}{\sim} B(1, \theta_2) \end{array}\right\}$$ with x_i & y_j independent for all i and j

Marking guidelines

2 pts Binary data
1 pt Independence within experimental conditions
1 pt Independence between experimental conditions
1 pt Parameter space. $\leq$ is $-\frac{1}{2}$. Full marks are possible even if (a) is a zero.

   (b) What is the parameter space for your model?

$$\{(\theta_1, \theta_2): 0 < \theta_1 < 1, 0 < \theta_2 < 1\}$$

2. (5 points) Homework Question 4 was about advertising awareness for Big Red chewing gum. In the last part of the question, you were asked to calculate the smallest n that yields a power of at least 0.9 when the true level of awareness was $\theta = 0.08$. **Write the required sample size (a number from your printout) in the space below. On your printout, circle the number and write "Question 2" beside it.**

$n = 1,653$

Please **fold your R printout into the quiz**, with your name on the quiz paper showing. Make sure your name and student number also appear on your printout. *If you do not have your R printout, do not answer Question 2.*

Name Jerry

Student Number

STA 442/2101 f2016 Quiz Two

1. (5 points) Let the test statistic T be continuous, with pdf $f(t)$ and cdf $F(t)$ under the null hypothesis.The null hypothesis is rejected if $T > c$. Show that if H_0 is true, the distribution of the p-value is $U(0,1)$. Derive the density. Start with the cumulative distribution function of the p-value: $Pr\{P \leq x\} = \ldots$. You may assume that $F(t)$ is strictly increasing over the support of the distribution, so it has a unique inverse.

For $0<x<1$,

$$Pr\{P \leq x\} = Pr\{1 - F(T) \leq x\} = Pr\{F(T) \geq 1-x\}$$

$$= Pr\{F^{-1}(F(T)) \geq F^{-1}(1-x)\} = Pr\{T \geq F^{-1}(1-x)\}$$

$$= 1 - F(F^{-1}(1-x)) = 1-(1-x) = x$$

(Full marks for saying this is the cdf of a uniform for $0<x<1$)

And the density of the p-value is

$$\frac{d}{dx} Pr\{P \leq x\} = \frac{d}{dx} x = 1 \text{ for } 0<x<1$$

This is the density of a $U(0,1)$.

2. (5 points) Homework Question 4 was about advertising awareness for Big Red chewing gum. In the last part of the question, you were asked to calculate the smallest n that yields a power of at least 0.9 when the true level of awareness was $\theta = 0.08$. **Write the required sample size (a number from your printout) in the space below. On your printout, circle the number and write "Question 2" beside it.**

$n = 1,653$

Please **fold your R printout into the quiz**, with your name on the quiz paper showing. Make sure your name and student number also appear on your printout. *If you do not have your R printout, do not answer Question 2.*

Name Jerry

Student Number

STA 442/2101 f2016 Quiz 3

1. (5 points) Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from a joint distribution with $E(X_i) = \mu_x$, $Var(X_i) = \sigma_x^2$, $E(Y_i) = \mu_y$, $Var(Y_i) = \sigma_y^2$ and $Cov(X_i, Y_i) = E(X_iY_i) - E(X_i)E(Y_i) = \sigma_{xy}$. A version of the sample covariance is

$$S_{xy} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}).$$

Prove that S_{xy} is a consistent estimator of σ_{xy}.

$$S_{xy} = \frac{1}{n}\sum_{i=1}^{n}(X_iY_i - X_i\overline{Y} - \overline{X}Y_i + \overline{X}\,\overline{Y})$$

$$= \frac{1}{n}\sum_{i=1}^{n}X_iY_i - \overline{X}\,\overline{Y} - \overline{X}\,\overline{Y} + \frac{1}{n}n\overline{X}\,\overline{Y} = \frac{1}{n}\sum_{i=1}^{n}X_iY_i - \overline{X}\,\overline{Y}$$

By LLN, $\frac{1}{n}\sum_{i=1}^{n}X_iY_i \xrightarrow{as} E(X_iY_i)$, $\overline{X} \xrightarrow{as} E(X_i)$ and $\overline{Y} \xrightarrow{as} E(Y_i)$

So by continuous mapping,

$$S_{xy} = \frac{1}{n}\sum_{i=1}^{n}X_iY_i - \overline{X}\,\overline{Y} \xrightarrow{as} E(X_iY_i) - E(X_i)E(Y_i)$$
$$= \sigma_{12}$$

2. (5 points) A chi-squared random variable X with parameter $\nu > 0$ has moment-generating function $M_X(t) = (1 - 2t)^{-\nu/2}$. Let $W = W_1 + W_2$, where W_1 and W_2 are independent, $W_2 \sim \chi^2(\nu_2)$ and $W \sim \chi^2(\nu_1 + \nu_2)$. The quantities ν_1 and ν_2 are both positive. Show $W_1 \sim \chi^2(\nu_1)$.

By independence

$$M_W(t) = M_{W_1}(t)\, M_{W_2}(t)$$

$$\Rightarrow (1-2t)^{-\frac{\nu_1+\nu_2}{2}} = M_{W_1}(t)\, \cancel{(1-2t)^{-\nu_2/2}}$$

$$\| $$

$$(1-2t)^{-\nu_1/2} \cancel{(1-2t)^{-\nu_2/2}}$$

$$\Rightarrow M_{W_1}(t) = (1-2t)^{-\nu_1/2} \quad , \text{ MGF of } \chi^2(\nu_1)$$

Name Jerry

Student Number

STA 442/2101 f2016 Quiz 4

1. (3 points) For the general linear model with fixed effects and normal errors (see formmula sheet) find the distribution of the vector of residuals **e**. Show your work and **circle your final answer.**

No marks off for not saying this

$e = (I-H)y = (I-H)(X\beta + \epsilon)$.

Since ϵ is multivariate normal, e is multivariate normal by the formula sheet.

$E(e) = (I-H)X\beta = X\beta - X\underbrace{(X^TX)^{-1}X^TX}_{I}\beta$

$= X\beta - X\beta = 0$ and

$cov(e) = cov((I-H)y) = (I-H)\,\sigma^2 I_n\,(I-H)^T$

$= \sigma^2 (I-H)$, so

$e \sim N(0, \sigma^2(I-H))$

No marks for starting with $e = y - \hat{y}$ and treating y and $\hat{y}$ as independent.

Name Jerry

Student Number

STA 442/2101 f2016 Quiz Four

1. (3 points) For the general linear model with fixed effects and normal errors (see formmula sheet) find the distribution of the vector of predicted values $\hat{\mathbf{y}}$. Show your work and **circle your final answer.**

No marks off for not saying this

$\hat{y} = X\hat{\beta}$. Since $\hat{\beta}$ is normal by the formula sheet, $\hat{y}$ is normal as well.

$$E(\hat{y}) = E(X\hat{\beta}) = XE(\hat{\beta}) = X\beta$$

$$cov(\hat{y}) = cov(X\hat{\beta}) = X\, cov(\hat{\beta}) X^T$$

$$= X \sigma^2 (X^T X)^{-1} X^T$$

Or they could start with $\hat{y} = Hy$. In either case,

$$\hat{y} \sim N(X\beta, \sigma^2 X(X^TX)^{-1}X^T)$$

either is okay

H

2. In a study comparing the effectiveness of different exercise programmes, volunteers were randomly assigned to one of three exercise programmes (A, B, C) or put on a waiting list and told to work out on their own. Aerobic capacity is the body's ability to process oxygen. Aerobic capacity was measured before and after 6 months of participation in the program (or 6 months of being on the waiting list). The response variable was improvement in aerobic capacity. The explanatory variables were age (a covariate) and treatment group. A regression model for this problem is

$$Y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 d_{i,1} + \beta_3 d_{i,2} + \beta_4 d_{i,3} + \epsilon_i,$$

where x_1 is age, and d_1, d_2 and d_3 are dummy variables for exercise program.

(a) (2 points) In the table below, indicate how the dummy variables are defined. There is more than one right answer, but make Wait list the reference category. In the last column, give the conditional expected improvement in terms of β values. The *symbols* for the dummy variables should not appear in the last column.

Programme	d_1	d_2	d_3	$E(y\|\mathbf{x})$
A	1	0	0	$\beta_0+\beta_2 + \beta_1 x$
B	0	1	0	$\beta_0+\beta_3 + \beta_1 x$
C	0	0	1	$\beta_0+\beta_4 + \beta_1 x$
Wait List	0	0	0	$\beta_0 + \beta_1 x$

(b) (1 point) What null hypothesis would you test to find out whether, controlling for age, the three exercise programmes differ in their average effectiveness? Give your answer in terms of β values.

$H_0: \beta_2 = \beta_3 = \beta_4 \cancel{= 0}$

(c) (1 point) What null hypothesis would you test to find out whether, controlling for age, the exercise programmes B is better than being on the wait list? Give your answer in terms of β values. In spite of the phrasing, it's a 2-tailed test.

$H_0: \beta_3 = 0$

(d) (1 point) What null hypothesis would you test to find out whether, controlling for exercise programme, age is related to improvement in aerobic capacity? Give your answer in terms of β values.

$H_0: \beta_1 = 0$

(e) (1 point) What is the difference in average benefit between programmes A and C for a 27 year old participant? Give your answer in terms of β values.

$\beta_2 - \beta_4$

(f) (1 point) Is it safe to assume that age is independent of exercise program? Answer Yes or No and indicate why. No marks without saying why.

Yes because of random assignment.

Name Jenny

Student Number

STA 442/2101 f2016 Quiz 5

1. (6 points) You have already proved that $(\mathbf{y}-X\beta)^\top(\mathbf{y}-X\beta) = \mathbf{e}^\top\mathbf{e}+(\widehat{\beta}-\beta)^\top(X^\top X)(\widehat{\beta}-\beta)$. Dividing both sides by σ^2, show that $\mathbf{e}^\top\mathbf{e}/\sigma^2 \sim \chi^2(n-p)$. Start with the distribution of the left side.

$$\frac{(y-X\beta)^T(y-X\beta)}{\sigma^2} = \frac{e^Te}{\sigma^2} + \frac{(\hat\beta-\beta)^T(X^TX)(\hat\beta-\beta)}{\sigma^2}$$

$$\Rightarrow (y-X\beta)^T(\sigma^2 I_n)^{-1}(y-X\beta) = \frac{e^Te}{\sigma^2} + (\hat\beta-\beta)^T(\sigma^2(X^TX)^{-1})^{-1}(\hat\beta-\beta)$$

$$W = W_1 + W_2$$

Because $y \sim N(X\beta, \sigma^2 I_n)$, $W \sim \chi^2(n)$

Because $\hat\beta \sim N(\beta, \sigma^2(X^TX)^{-1})$, $W_2 \sim \chi^2(p)$

Because $\hat\beta$ & e are independent, W_1 & W_2 are indepen-dent

So by the formula sheet, $W_1 \sim \chi^2(n-p)$

TD

Name Jerry

Student Number __________

STA 442/2101 f2016 Quiz Five

1. (6 points) Assuming the general linear model with normal errors (see formula sheet), let $\mathbf{a}$ be a $p \times 1$ vector of constants.

   (a) What is the distribution of $\mathbf{a}^\top \widehat{\boldsymbol{\beta}}$? Your answer includes both the expected value and the variance.

   $$a^T\hat{\beta} \sim N(a^T\beta, \sigma^2 a^T(X^TX)^{-1}a)$$

   (b) Now standardize the difference (subtract off the mean and divide by the standard deviation) to obtain a standard normal.

   $$Z = \frac{a^T\hat{\beta} - a^T\beta}{\sqrt{\sigma^2 a^T(X^TX)^{-1}a}}$$

   (c) Divide by the square root of a well-chosen chi-squared random variable, divided by its degrees of freedom, and simplify. Call the result T.

   $$T = \frac{(a^T\hat{\beta} - a^T\beta)/\sqrt{\sigma^2 a^T(X^TX)^{-1}a}}{\sqrt{\frac{e^Te}{\sigma^2}/(n-p)}}$$

   $$= \frac{a^T\hat{\beta} - a^T\beta}{\sqrt{MSE\, a^T(X^TX)^{-1}a}}$$

   (d) How do you know numerator and denominator are independent?

   Because $\hat{\beta}$ & e are independent

Name Jerry

Student Number

STA 442/2101 f2016 Quiz 6

1. (5 points) For the general linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, consider the null hypothesis $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{h}$. Let $\mathbf{A}$ be an $r \times r$ nonsingular matrix (meaning $\mathbf{A}^{-1}$ exists), so that $\mathbf{L}\boldsymbol{\beta} = \mathbf{h}$ if and only if $\mathbf{AL}\boldsymbol{\beta} = \mathbf{Ah}$. Show that the general linear test statistic F for testing $H_0 : (\mathbf{AL})\boldsymbol{\beta} = \mathbf{Ah}$ is the same as the one for testing $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{h}$.

$$F^* = \frac{(AL\hat{\beta} - Ah)^T (AL(X^TX)^{-1}(AL)^T)^{-1} (AL\hat{\beta} - Ah)}{r\, MSE}$$

$$= \frac{(A(L\hat{\beta} - h))^T (AL(X^TX)^{-1}L^TA^T)^{-1} A(L\hat{\beta} - h)}{r\, MSE}$$

$$= \frac{(L\hat{\beta} - h)^T A^T (A L(X^TX)^{-1}L^TA^T)^{-1} A(L\hat{\beta} - h)}{r\, MSE}$$

$$= \frac{(L\beta - h)^T \underbrace{A^T A^{T^{-1}}}_{I} (L(X^TX)^{-1}L^T)^{-1} \underbrace{A^{-1}A}_{I}(L\hat{\beta} - h)}{r\, MSE}$$

$$= \frac{(L\hat{\beta} - h)^T (L(X^TX)^{-1}L^T)^{-1} (L\hat{\beta} - h)}{r\, MSE}$$

Test statistic for $H_0: L\beta = h$

Name Jerry

Student Number

STA 442/2101 f2016 Quiz 6

1. (5 points) For the general linear model, consider a one-to-one transformation of the explanatory variables by $\mathbf{X}^* = \mathbf{XA}$, where $\mathbf{A}$ is a $p \times p$ matrix with an inverse. Suppose we use the transformed explanatory variables instead of the original ones, and fit the model $\mathbf{y} = \mathbf{X}^*\beta + \epsilon$. Show that the vector of predicted values $\widehat{\mathbf{y}}^* = \mathbf{X}^*\widehat{\beta}^*$ is equal to the usual $\widehat{\mathbf{y}}$, so that predictions are unaffected.

$$\hat{\beta}^* = (X^{*T}X^*)^{-1}X^{*T}y = ((XA)^T XA)^{-1}(XA)^T y$$

$$= (A^T X^T X A)^{-1} A^T X^T y$$

$$= A^{-1}(X^T X)^{-1}\underbrace{A^{T^{-1}} A^T}_{I} X^T y$$

$$= A^{-1}\hat{\beta}, \text{ so}$$

$$\hat{y}^* = X^*\hat{\beta}^* = X\underbrace{AA^{-1}}_{I}\hat{\beta} = X\hat{\beta} = \hat{y}$$

Name Jerry

Student Number

STA 442/2101 f2016 Quiz 7

1. Suppose you fit (estimate the parameters of) a regression model, obtaining $\widehat{\boldsymbol{\beta}}$, $\widehat{\mathbf{y}}$ and $\mathbf{e}$. Call this the *first model.* Then as an experiment, you fit a second regression model, using $\widehat{\mathbf{y}}$ from the first model as the response variable, and exactly the same $\mathbf{X}$ matrix as the first model. Call this the *second model.* **The following questions are pretty easy and you have more room than you need. Don't over-think this.**

   (a) (2 points) What is $\widehat{\boldsymbol{\beta}}$ for the second model? Denote it by $\widehat{\boldsymbol{\beta}}_2$.

   $$\hat{\beta}_2 = (X^T X)^{-1} X^T \hat{y} = (X^T X)^{-1} X^T X \hat{\beta} = \hat{\beta}$$

   (b) (2 points) What is $\widehat{\mathbf{y}}$ for the second model? Denote it by $\widehat{\mathbf{y}}_2$.

   $$\hat{y}_2 = X \hat{\beta}_2 = X \beta = \hat{y}$$

   (c) (2 points) What is $\mathbf{e}$ for the second model? Denote it by $\mathbf{e}_2$.

   $$e_2 = \hat{y} - \hat{y}_2 = \hat{y} - \hat{y} = 0$$

Name Jerry

Student Number

STA 442/2101 f2016 Quiz 7

1. Suppose you fit (estimate the parameters of) a regression model, obtaining $\widehat{\boldsymbol{\beta}}$, $\widehat{\mathbf{y}}$ and $\mathbf{e}$. Call this the *first model*. Then as an experiment, you fit a second regression model, using $\mathbf{e}$ from the first model as the response variable, and exactly the same $\mathbf{X}$ matrix as the first model. Call this the *second model*. **The following questions are pretty easy and you have more room than you need. Don't over-think this.**

   (a) (2 points) What is $\boldsymbol{\beta}$ for the second model? Denote it by $\widehat{\boldsymbol{\beta}}_2$. Show some work and simplify.

   $$\hat{\beta}_2 = (X^TX)^{-1}\underbrace{X^Te}_{0} = 0$$

   no marks off for not showing this

   (b) (2 points) What is $\widehat{\mathbf{y}}$ for the second model? Denote it by $\widehat{\mathbf{y}}_2$. Show some work and simplify.

   $$\hat{y}_2 = X\hat{\beta}_2 = X0 = 0$$

   (c) (2 points) What is $\mathbf{e}$ for the second model? Denote it by $\mathbf{e}_2$. Show some work and simplify.

   $$e_2 = e - \hat{y}_2 = e - 0 = e$$

Name Jerry

Student Number

STA 442/2101 f2016 Quiz 8

(10 points) One version of the delta method says that if $X_1, \ldots, X_n$ are a random sample from a distribution with mean μ and variance σ^2, and $g(x)$ is a function whose derivative is continuous in a neighbourhood of $x = \mu$, then $\sqrt{n}\left(g(\overline{X}_n) - g(\mu)\right) \stackrel{d}{\rightarrow} T \sim N(0, g'(\mu)^2\sigma^2)$. In many applications, both μ and σ^2 are functions of some parameter θ.

Let $X_1, \ldots, X_n$ be a random sample from an exponential distribution with parameter θ, so that $E(X_i) = \theta$ and $Var(X_i) = \theta^2$. Find a function $g(x)$ such that the limiting distribution of $Z_n = \sqrt{n}\left(g(\overline{X}_n) - g(\theta)\right)$ is *standard* normal — that is $Z_n \stackrel{d}{\rightarrow} Z \sim N(0, 1)$. Show your work. Finish your answer with the words "The function is ..." Write the function and **circle it**.

$$g'(\theta)^2 \cdot \theta^2 = 1 \iff g'(\theta)^2 = \frac{1}{\theta^2}$$

$$\Rightarrow g'(\theta) = \frac{1}{\theta}$$ (Just assume $g'(\theta) > 0$. No need to mention it.)

That is, $\frac{dg}{d\theta} = \frac{1}{\theta} \Rightarrow dg = \frac{1}{\theta}\, d\theta$

$$\Rightarrow \int dg = \int \frac{1}{\theta}\, d\theta = \ln(\theta) + c$$

let it $= 0$

So the function is $g(\theta) = \ln(\theta)$

They don't have to use separation of variables. They can even guess if they guess right.

Name Jerry

Student Number

STA 442/2101 f2016 Quiz 8

(10 points) One version of the delta method says that if $X_1, \ldots, X_n$ are a random sample from a distribution with mean μ and variance σ^2, and $g(x)$ is a function whose derivative is continuous in a neighbourhood of $x = \mu$, then $\sqrt{n}\left(g(\overline{X}_n) - g(\mu)\right) \stackrel{d}{\rightarrow} T \sim N(0, g'(\mu)^2\sigma^2)$. In many applications, both μ and σ^2 are functions of some parameter θ.

Let $X_1, \ldots, X_n$ be a random sample from a Poisson distribution with parameter λ, so that $E(X_i) = \lambda$ and $Var(X_i) = \lambda$. Find a function $g(x)$ such that the limiting distribution of $Z_n = \sqrt{n}\left(g(\overline{X}_n) - g(\lambda)\right)$ is *standard* normal — that is $Z_n \stackrel{d}{\rightarrow} Z \sim N(0,1)$. Show your work. Finish your answer with the words "The function is ..." Write the function and **circle it**.

$$g'(\lambda)^2 \cdot \lambda = 1 \Rightarrow g'(\lambda)^2 = \frac{1}{\lambda}$$

$$\Rightarrow g'(\lambda) = \lambda^{-\frac{1}{2}}$$ (Just assume $g'(\lambda) > 0$ No need to mention it.)

That is $$\frac{dg}{d\lambda} = \lambda^{-\frac{1}{2}} \Rightarrow dg = \lambda^{-\frac{1}{2}} d\lambda$$

$$\Rightarrow \int dg = \int \lambda^{-\frac{1}{2}} d\lambda = \frac{\lambda^{1/2}}{1/2} + \underset{\uparrow \text{ Let it} = 0}{c} = 2\sqrt{\lambda}$$

So The function is $g(\lambda) = 2\sqrt{\lambda}$

They don't have to use separation of variables. They can even guess if they guess right.

Name Jerry

Student Number

STA 442/2101 f2017 Quiz 1

1. (5 points) The $n \times n$ real matrix $\mathbf{A}$ has an inverse, but it is not necessarily symmetric, so the spectral decomposition does *not* apply. Recall that $\mathbf{A}$ has n eigenvalue-eigenvector pairs, satisfying $\mathbf{A}\mathbf{x}_j = \lambda_j \mathbf{x}_j$ for $j = 1, \ldots, n$, where $\mathbf{x}_j^\top \mathbf{x}_j = 1$. Because $\mathbf{A}^{-1}$ exists, the eigenvalues of $\mathbf{A}$ are all non-zero, something you do not have to prove.

   In terms of $\lambda_1, \ldots, \lambda_n$, what are the eigenvalues of $\mathbf{A}^{-1}$? Show the calculation and then state the answer. **Circle your answer**. You have more room than you need.

For $j=1,\ldots,n$, $A x_j = \lambda_j x_j$

$\Rightarrow A^{-1} A x_j = A^{-1} \lambda_j x_j = \lambda_j A^{-1} x_j$

$\Rightarrow x_j = \lambda_j A^{-1} x_j$

$\Rightarrow A^{-1} x_j = \left(\frac{1}{\lambda_j}\right) x_j$

The eigenvalues of A^{-1} are one over the eigenvalues of A.

2. (5 points) Let $X_1, \ldots, X_n$ be a random sample (meaning independent and identically distributed) from a distribution with density $f(x|\theta) = \theta x^{\theta-1}$ for $0 < x < 1$, where the parameter $\theta > 0$. Find the maximum likelihood estimate of θ. Show your work. Don't bother with the second derivative test. Your final answer is a formula for computing $\widehat{\theta}$ from $X_1, \ldots, X_n$. **Circle it.**

$$L(\theta) = \prod_{i=1}^{n} \theta x_i^{\theta-1} = \theta^n \left(\prod_{i=1}^{n} x_i\right)^{\theta-1}$$

$$\ell(\theta) = \ln L(\theta) = n \ln \theta + (\theta - 1) \sum_{i=1}^{n} \ln x_i$$

$$\ell'(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} \ln x_i \overset{set}{=} 0$$

$$\Rightarrow \frac{n}{\theta} = -\sum_{i=1}^{n} \ln x_i \Rightarrow \theta = \frac{-n}{\sum_{i=1}^{n} \ln x_i} \text{ and}$$

$$\hat{\theta} = -\frac{n}{\sum_{i=1}^{n} \ln x_i}$$

At most one point for any answer that could not be computed from sample data.

No distinction between estimator & estimate here.

Name Jerry

Student Number

STA 442/2101 f2017 Quiz One

1. (5 points) The $n \times n$ real matrix $\mathbf{A}$ is not necessarily symmetric, so the spectral decomposition does *not* apply. Recall that $\mathbf{A}$ has n eigenvalue-eigenvector pairs, satisfying $\mathbf{A}\mathbf{x}_j = \lambda_j \mathbf{x}_j$ for $j = 1, \ldots, n$, where $\mathbf{x}_j^\top \mathbf{x}_j = 1$.

   In terms of $\lambda_1, \ldots, \lambda_n$, what are the eigenvalues of $\mathbf{A}^2 = \mathbf{A}\mathbf{A}$? Show the calculation and then state the answer. **Circle your answer**. You have more room than you need.

For $j = 1, \ldots, n$ $\quad A x_j = \lambda_j x_j$

$\Longrightarrow AA x_j = A \lambda_j x_j = \lambda_j A x_j = \lambda_j \lambda_j x_j$

$\Longrightarrow A^2 x_j = \lambda_j^2 x_j$

The eigenvalues of A^2 are the eigenvalues of A, squared.

2. (5 points) Let $X_1, \ldots, X_n$ be a random sample (meaning independent and identically distributed) from a distribution with density $f(x|\theta) = \theta(1-x)^{\theta-1}$ for $0 < x < 1$, where the parameter $\theta > 0$. Find the maximum likelihood estimate of θ. Show your work. Don't bother with the second derivative test. Your final answer is a formula for computing $\widehat{\theta}$ from $X_1, \ldots, X_n$. **Circle it.**

$$L(\theta) = \prod_{i=1}^{n} \theta(1-x_i)^{\theta-1} = \theta^n \left(\prod_{i=1}^{n}(1-x_i)\right)^{\theta-1}$$

$$\ell(\theta) = \ln L(\theta) = n \ln \theta + (\theta-1)\sum_{i=1}^{n} \ln(1-x_i)$$

$$\ell'(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} \ln(1-x_i) \overset{set}{=} 0$$

$$\Rightarrow \frac{n}{\theta} = -\sum_{i=1}^{n} \ln(1-x_i) \Rightarrow \theta = \frac{-n}{\sum_{i=1}^{n} \ln(1-x_i)}$$

And

$$\hat{\theta} = -\frac{n}{\sum_{i=1}^{n} \ln(1-x_i)}$$

<u>At most</u> one point for any answer that could not be computed from sample data.

No distinction between estimator & estimate here.

3. (6 points) Mantids are insects, like crickets or grasshoppers. When frightened, they emit loud noises that function as alarm calls. The frequency (number of calls per minute) may indicate how alarmed the mantids are. In one study, caged mantids were randomly assigned to be exposed to one of four predators (birds), and the number of alarm calls per minute was recorded.

   (a) Write a regression equation ($Y_i = \ldots$ and so on) in which expected number of alarm calls is a function of Predator. You will use indicator dummy variable coding *with* an intercept, and an indicator for each treatment except the reference category. Just give the regression equation in the space below.

   $$y_i = \beta_0 + \beta_1 P_{i1} + \beta_2 P_{i2} + \beta_3 P_{i3} + \varepsilon_i$$

   (b) Make a table with one row for each predator, showing how the dummy variables are defined. Label the 4 predators A, B, C and D. Make one more column with $E(Y|\mathbf{x})$. *Symbols* for the dummy variables must not appear in the $E(Y|\mathbf{x})$ column.

Predator	P_1	P_2	P_3	$E(y\|x)$
A	1	0	0	$\beta_0 + \beta_1$
B	0	1	0	$\beta_0 + \beta_2$
C	0	0	1	$\beta_0 + \beta_3$
D	0	0	0	β_0

   (c) In terms of the β values from your regression model, give the null hypothesis you would test to answer each of the following questions.

      i. Are there *any* differences among predators in the expected frequency of alarm calls?

      $H_0: \beta_1 = \beta_2 = \beta_3 = 0$

      ii. Do predators A and B differ in the average number of alarm calls they elicit?

      $H_0: \beta_1 = \beta_2$

      iii. Do predators A and D differ in the average number of alarm calls they elicit?

      $H_0 \; \beta_1 = 0$

Please attach your printout to the quiz. Fold it so that your name on the quiz is showing.

I don't have my printout. My dog ate it but
I remember all the numbers.

Name Jerry

Student Number

STA 442/2101 f2017 Quiz 4

(10 points) In this question you may use the (strong) consistency of the sample variance $\frac{1}{n}\sum_{i=1}^n(X_i-\overline{X})^2$ and the sample covariance $\frac{1}{n}\sum_{i=1}^n(X_i-\overline{X})(Y_i-\overline{Y})$ without proof, as well as handy formulas that are not on the formula sheet, like $Cov(X,Y) = E(XY)-E(X)E(Y)$.

Independently for $i = 1,\ldots,n$, let $Y_i = \beta X_i + \epsilon_i$, where $E(X_i) = \mu$, $E(\epsilon_i) = 0$, $Var(X_i) = \sigma^2_x$, $Var(\epsilon_i) = \sigma^2_\epsilon$, and ϵ_i is independent of X_i. Let

$$\widehat{\beta}_n = \frac{\sum_{i=1}^n(X_i-\overline{X})(Y_i-\overline{Y})}{\sum_{i=1}^n(X_i-\overline{X})^2}.$$

Is $\widehat{\beta}_n$ a consistent estimator of β? Answer Yes or No and prove your answer. At the end of your calculations, write *Yes, consistent* or *No, not consistent.*

$$Cov(X_i, Y_i) = E(X_iY_i) - E(X_i)E(Y_i) = E\{X_i(\beta X_i + \varepsilon_i)\} - \mu\beta\mu$$

$$= \beta E(X_i^2) + E(X_i)E(\varepsilon_i) - \beta\mu^2$$

$$= \beta(E(X_i^2) - \mu^2) = \beta\sigma_x^2, \text{ so}$$

$$\hat{\beta}_n = \frac{\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X})(Y_i-\bar{Y})}{\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X})^2} \xrightarrow{as} \frac{Cov(X_i,Y_i)}{Var(X_i)}$$

By SLLN and continuous mappings

$$= \frac{\beta\sigma_x^2}{\sigma_x^2} = \beta$$

Yes, consistent

Name ______________________

Student Number ______________________

STA 442/2101 f2017 Quiz 5

1. (5 points) Let $X_1, \ldots, X_n$ independent and identically distributed random variables from a distribution with density $f(x|\theta) = \frac{1}{\theta} e^{-x/\theta}$ for $x > 0$, where $\theta > 0$. This is one way to parameterize the exponential distribution, so $E(X_i) = \theta$ and $Var(X_i) = \theta^2$. Using the delta method, find a function $g(x)$ such that

$$\sqrt{n}\left(g(\overline{X}_n) - g(\mu)\right) \stackrel{d}{\rightarrow} Z \sim N(0,1)$$

Show your work. The final answer is a formula for $g(x)$. **Circle the formula.**

CLT says $\sqrt{n}(\overline{X}_n - \theta) \xrightarrow{d} T \sim N(0, \frac{1}{\theta^2})$

Delta says $\sqrt{n}(g(\bar{X}_n) - g(\mu)) \xrightarrow{d} g'(\mu) T \sim N(0, g'(\mu)^2 \sigma^2)$

So want $g'(\theta)^2 \cdot \frac{1}{\theta^2} = 1$, for example $g'(\theta) = \theta$

Try $g(x) = \frac{1}{2}x^2$ so $g'(x) = \frac{1}{2} \cdot 2x = x$,

and $\sqrt{n}\left(\frac{1}{2}\bar{X}_n^2 - \frac{1}{2}\theta^2\right) \to Z = \theta T \sim N(0,1)$

$g(x) = \frac{1}{2}x^2$

2. (5 points) Recall that if $X \sim \chi^2(\nu)$, it has moment-generating function $M_X(t) = (1-2t)^{-\nu/2}$. Let $W = W_1 + W_2$, where W_1 and W_2 are independent, $W \sim \chi^2(\nu_1 + \nu_2)$ and $W_2 \sim \chi^2(\nu_2)$, ~~where~~ with ν_1 and ν_2 ~~are~~ both positive. Show $W_1 \sim \chi^2(\nu_1)$. You may use the fact that the moment-generating function of a sum of independent random variables is the product of moment-generating functions.

$M_W(t) = M_{W_1}(t)\, M_{W_2}(t)$, so

$$(1-2t)^{-\frac{\nu_1+\nu_2}{2}} = M_{W_1}(t)\,(1-2t)^{-\nu_2/2}$$

$$\Rightarrow (1-2t)^{-\nu_1/2}\,\cancel{(1-2t)^{-\nu_2/2}} = M_{W_1}(t)\,\cancel{(1-2t)^{-\nu_2/2}}$$

$$\Rightarrow M_{W_1}(t) = (1-2t)^{-\nu_1/2}$$

MGF of $\chi^2(\nu_1)$

Name ____________________

Student Number ____________________

STA 442/2101 f2017 Quiz Five

1. (5 points) Let $X_1, \ldots, X_n$ independent and identically distributed random variables from a distribution with density $f(x|\theta) = \frac{1}{\Gamma(\theta)} e^{-x} x^{\theta-1}$ for $x > 0$, where $\theta > 0$. This is a version of the gamma distribution, with $E(X_i) = Var(X_i) = \theta$. Find a variance-stabilizing transformation for the sample mean. Specifically, find a function $g(x)$ such that
$$\sqrt{n}\left(g(\overline{X}_n) - g(\mu)\right) \stackrel{d}{\rightarrow} Z \sim N(0,1)$$
Show your work. The final answer is a formula for $g(x)$. **Circle the formula.**

CLT says $\sqrt{n}(\overline{X}_n - \theta) \xrightarrow{d} T \sim N(0, \theta)$.

Delta says $\sqrt{n}\left(g(\overline{X}_n) - g(\theta)\right) \xrightarrow{d} g'(\theta) T \sim N(0, g'(\theta)^2 \theta)$

So want $g'(\theta)^2 \theta = 1$ or $|g'(\theta)| = \frac{1}{\sqrt{\theta}} = \theta^{-\frac{1}{2}}$

So try $g(x) = 2\sqrt{x} = 2x^{\frac{1}{2}}$, and

$g'(x) = 2 \cdot \frac{1}{2} x^{-\frac{1}{2}} = \frac{1}{\sqrt{\theta}}$, and

$\sqrt{n}\left(2\sqrt{\overline{X}_n} - 2\sqrt{\theta}\right) \xrightarrow{d} \frac{1}{\sqrt{\theta}} T \sim N(0,1)$

$g(x) = 2\sqrt{x}$

2. (5 points) Let the $p \times 1$ random vector $\mathbf{w}$ have mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$; let $\mathbf{A}$ be a $q \times p$ matrix of constants and let $\mathbf{B}$ be an $r \times p$ matrix of constants. Derive a simple formula for $cov(\mathbf{Aw}, \mathbf{Bw})$. **Circle the formula.**

$$cov(Aw, Bw) = E\{(Aw - A\mu)(Bw - B\mu)^T\}$$

$$= E\{A(w-\mu)(B(w-\mu))^T\}$$

$$= E\{A(w-\mu)(w-\mu)^T B^T\}$$

$$= A\, E\{(w-\mu)(w-\mu)^T\} B^T$$

$$= A\, cov(w)\, B^T = A \Sigma B^T$$