Previous exams MW241

Oliver Kirchkamp

9th October 2019

Contents

1 Instructions for the exam

- In the exam you obtain the following material:
  - Answer sheet (Please make sure you have obtained the correct answer sheet with your name.)
  - Instructions
  - Scratch paper
- At the end of the exam please submit *only the answer sheet.* You will keep the task sheet, your scratch paper and the instructions. If you mark your answers on the task sheet, you can easily compare your answers with the solutions to the exam.

  Answer sheets will be collected in the same order in which task sheet were distributed. Thus, everybody has approximately the same time to work on the exam. To make sure that nobody has an undue advantage, **late submissions do not count.**

  Please remain *quietly at your place, until the last person has completed the exam.* All participants should have the chance to complete their exam in a quiet environment.
- You need *only paper and pen* for this exam. If necessary, please discuss eventual needs for special support before the exam. Bags, calculators, mobile phones, books, etc. are left at the front of the room. Please make sure you have completely switched off your mobile devices before you leave them.
- There are different version of the exam. Your version code is indicated at the top of the task sheet, as follows:

  ☐☐☐☐☐

  Please copy your version code from the task sheet into your answer sheet.
- Please mark your answers with a dark pen like this:

answer	a	b	c	d	e
x.x	☐	☐	☐	☒	☐

  You only score points if you tick the correct answer. You obtain no points if you tick the wrong answer.

Some questions may have more than one correct answer. These questions are denoted with the text "multiple answers possible". For each correct "yes" and for each correct "no" you obtain a fraction of the points for this question.
Example: There are five possible answers: a, b, c, d, and e. Correct are c, d and e. You score all points with the following answer:

answer	a	b	c	d	e
x.x: yes	☐	☐	☒	☒	☒
x.x: no	☒	☒	☐	☐	☐

You score 4/5 of the points with the following:

answer	a	b	c	d	e
x.x: yes	☐	☐	☐	☒	☒
x.x: no	☒	☒	☒	☐	☐

(a,b,d,e is answered correctly, c is not correct). For the following answer you obtain 2/5 of the points:

answer	a	b	c	d	e
x.x: yes	☐	☒	☐	☒	☐
x.x: no	☒	☐	☒	☐	☒

(only a and d are correct). For the following answer you obtain no points:

answer	a	b	c	d	e
x.x: yes	☒	☒	☐	☐	☐
x.x: no	☐	☐	☒	☒	☒

(no answer is correct).
Please always indicate "yes" or "no". Otherwise your exam has to be graded manually. This slows down the process for you and

for everybody else. Please note that you get already some points by giving random answers. To pass you will need more points than what you can get with random answers.

The last five minutes:

- Please copy your answers into the answer sheet **only when you are sure about your answers.**

  If you, nevertheless, have to "clear" a field, please cross out the answer as in this example. Here only (b) is marked as correct.

answer	a	b	c	d	e
x.x:	☐	☒	☐	✱	☐

- Please do not forget to sign your answer sheet with indelible ink (e.g. a ballpoint pen).
- We wish you good luck

Distributions in R and JAGS:

type	R distribution (F)	R quantile (Q)	JAGS density
normal distribution	pnorm	qnorm	dnorm
t-distribution	pt	qt	dt
χ^2-distribution	pchisq	qchisq	dchisqr
F-distribution	pf	qf	df
Γ-distribution	pgamma	qgamma	dgamma
Poisson-distribution	ppois	qpois	dpois
NB-distribution	pnbinom	qnbinom	dnegbin

Poisson distribution: $P_\lambda(X=k) = \lambda^k \cdot e^{-\lambda}/k!$; $E[X]=\lambda$; $\mathrm{var}(X)=\lambda$

Negative binomial distribution: $X \sim \mathrm{NB}(\mu,\theta) = \mathrm{NB}'(p,\theta), p = \theta/(\theta+\mu)$, $E[X]=\mu, \mathrm{var}(X) = \mu + \mu^2/\theta$.

Exponential distribution: $f_\lambda(X) = \begin{cases} \lambda e^{-\lambda X} & X \geq 0 \\ 0 & \text{otherwise} \end{cases}$;

$F_\lambda(X) = \begin{cases} 1-e^{-\lambda X} & X \geq 0 \\ 0 & \text{otherwise} \end{cases}$;

$E[X] = 1/\lambda$; $\mathrm{var}(X) = 1/\lambda^2$

Gamma distribution: $X \sim \Gamma(\alpha,\beta)$, $E[X]=\alpha/\beta$, $\mathrm{var}(X) = \alpha/\beta^2$.

Some integrals: $\int x\,dx = \frac{1}{2}x^2 + C$; $\int x^n\,dx = x^{n+1}/(n+1) + C$; $\int \frac{1}{x}\,dx = \ln x + C$; $\int a^x\,dx = a^x/\ln a + C$

Derivative of the Log-Likelihood function:

$$\frac{d}{d\theta}\ln L(x_1,\ldots,x_n|\theta) = \frac{f'(x_1|\theta)}{f(x_1|\theta)} + \cdots + \frac{f'(x_n|\theta)}{f(x_n|\theta)}$$

Expected value: $E(c\cdot X) = c\cdot E(X)$; $E(X+Y) = E(X)+E(Y)$

Variance: $\mathrm{var}(c\cdot X) = c^2\cdot\mathrm{var}(X)$; $\mathrm{var}(X+Y) = \mathrm{var}(X)+\mathrm{var}(Y) + 2\cdot cov(X,Y)$

Variance of $\bar{X}$: $\mathrm{var}(\bar{X}) = \sigma_X^2/n$

Standard deviation of $\bar{X}$: $\sigma_{\bar{X}} = \sigma_X/\sqrt{n}$

Estimator for expected value: $\hat{\mu}_X = \bar{X} = \frac{1}{n}\sum_i X_i$

Estimator for variance: $\hat{\sigma}_X^2 = \frac{1}{n-1}\sum_{i=1}^n \left(X_i - \bar{X}\right)^2$

Estimator for standard deviation of X:

$\hat{\sigma}_X = \sqrt{\frac{1}{n-1}\sum_{i=1}^n \left(X_i-\bar{X}\right)^2}$

Estimator for $\sigma_{\bar{X}}$: $\hat{\sigma}_{\bar{X}} = \hat{\sigma}_X/\sqrt{n}$

Bias: $\mathrm{Bias}(\hat{\theta},\theta) = E(\hat{\theta}) - \theta$

Confidence interval for the mean: $\left[\bar{X} + \sigma_{\bar{X}}\cdot Q\left(\frac{\alpha}{2}\right); \bar{X} - \sigma_{\bar{X}}\cdot Q\left(\frac{\alpha}{2}\right)\right]$

Type I and II error:

test result	actual condition: H_0 false	actual condition: H_0 true
reject H_0 (positive)	$1-\beta$, Power sensitivity	α, significance type I err.
do not reject H_0 (negative)	β type II err.	$1-\alpha$ specificity

Comparing means (independent samples)

$$\frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{\hat{\sigma}_A^2}{n_A} + \frac{\hat{\sigma}_B^2}{n_B}}} \sim t_{n_A+n_B-2}$$

Comparing means (paired samples)

$g = \frac{\bar{\Delta}}{\hat{\sigma}_{\bar{\Delta}}} \sim t_{n-1}$ mit $\Delta_i = X_i - Y_i$ und

$\hat{\sigma}_{\bar{\Delta}} = \frac{1}{\sqrt{n}}\sqrt{\frac{\sum_{i=1}^n(\Delta_i-\bar{\Delta})^2}{n-1}}$

χ^2-independence test $e_{ij} = \frac{\sum_{j=1}^k X_{ij}\cdot\sum_{i=1}^n X_{ij}}{\sum_{i=1}^n\sum_{j=1}^k X_{ij}}$

$g = \sum_{i=1}^n\sum_{j=1}^k \frac{\left(X_{ij}-e_{ij}\right)^2}{e_{ij}} \sim \chi^2_{(n-1)\cdot(k-1)}$

χ^2-goodness-of-fit test: $g = \sum_{i=1}^k \frac{\left(X(a_i) - n\cdot P(a_i)\right)^2}{n\cdot P(a_i)} \sim \chi^2_{k-1}$

Testing means: $g = \frac{\bar{X}-\mu_0}{\hat{\sigma}_{\bar{X}}}$. If $X \sim N : g \sim t_{n-1}$ or $g^2 \sim F_{1,n-1}$.

If $n\to\infty : g \sim N(0,1)$ or $g^2 \sim F_{1,\infty}$.

AIC $= -2\cdot L + 2\cdot k$ (L is the Likelihood of the model, k is the number of parameters).

Logistic function: $\mathcal{L}(x) = 1/(1+e^{-x})$

$\mathrm{odds}(x) = \mathcal{L}(x)/(1-\mathcal{L}(x)) = e^x$

Precision: $\tau = 1/\sigma^2$

Conjugate priors:

$X \sim N(\mu,\tau), \mu \sim N(\mu_0,\tau_0)$:

$$\mu_{\text{post}} = \frac{\tau_0\mu_0 + n\tau\bar{x}}{\tau_0 + n\tau}$$

$$\tau_{\text{post}} = \tau_0 + n\tau$$

$X \sim N(\mu,\tau), \tau \sim \Gamma(\alpha_0,\beta_0)$:

$$\text{shape} \quad \alpha_{\text{post}} = \alpha_0 + \frac{n}{2}$$

$$\text{rate} \quad \beta_{\text{post}} = \beta_0 + \frac{n}{2}\mathrm{var}(x)$$

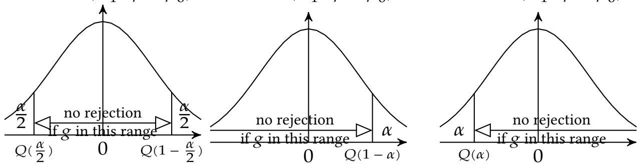
$X \sim \mathrm{bern}(p), p \sim \mathrm{Beta}(\alpha_0,\beta_0)$

$$\alpha_{\text{post}} = \alpha + \sum x_i$$

$$\beta_{\text{post}} = \beta + n - \sum x_i$$

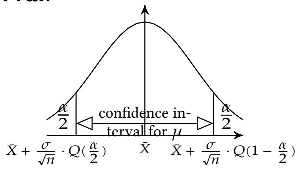
Test of significance:

test statistic: $g = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$



H_0 is rejected if g is outside the range.

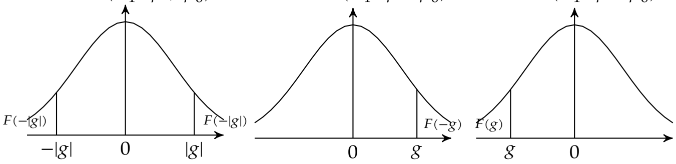
Confidence interval:



$H_0 : \mu = \mu_0$ is rejected if μ_0 is outside the confidence interval for μ.

p**-value:** test statistic: $g = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$

two sided ($H_1 : \mu \neq \mu_0$) | one sided ($H_1 : \mu > \mu_0$) | one sided ($H_1 : \mu < \mu_0$)



$p = 2 \cdot F(-|g|)$ | $p = F(-g)$ | $p = F(g)$

H_0 is rejected if $p < \alpha$.

	0.9	0.95	0.975	0.99	0.995	0.9975	0.999
qnorm(x)	1.28	1.64	1.96	2.33	2.58	2.81	3.09
qt(x,1)	3.08	6.31	12.71	31.82	63.66	127.32	318.31
qt(x,2)	1.89	2.92	4.30	6.96	9.92	14.09	22.33
qt(x,3)	1.64	2.35	3.18	4.54	5.84	7.45	10.21
qt(x,4)	1.53	2.13	2.78	3.75	4.60	5.60	7.17
qt(x,5)	1.48	2.02	2.57	3.36	4.03	4.77	5.89
qt(x,6)	1.44	1.94	2.45	3.14	3.71	4.32	5.21
qt(x,7)	1.41	1.89	2.36	3.00	3.50	4.03	4.79
qt(x,8)	1.40	1.86	2.31	2.90	3.36	3.83	4.50
qt(x,9)	1.38	1.83	2.26	2.82	3.25	3.69	4.30
qt(x,10)	1.37	1.81	2.23	2.76	3.17	3.58	4.14

2 MW24.1, April 2019 Final Resit

Question 1: You have estimated in R an OLS model with the command `est <- lm(Y ~ X)`. You assume that residuals follow a normal distribution. You want to use a parametric bootstrap to estimate a standard deviation of the estimated coefficient of `X`. To do this, you issue the following command (In the following `N` denotes the sample size, α is a placeholder):

```
sd(replicate(10000,{ X<-α; coef(lm(Y ~ X))['X']}))
```

What can you use for α? (3 points)

1a: None of the following answers are correct.

1b: `sample(est)+residuals(est))`

1c: `sample(est)+rnorm(N,sd=sd(residuals(est)))`

1d: `predict(est)+rnorm(N,sd=sd(sample(est)))`

1e: `predict(est)+rnorm(N,sd=sd(residuals(est)))`

Question 2: You estimate two coefficients, β_0 and β_x, as follows:

```
est <- glm(y ~ x,family=gaussian)
```

For which alternative hypothesis does the following expression approximate a p-value?

```
B<-replicate(10000,{
  est <- glm(y ~ sample(x),family=gaussian)
  coef(est)[2] })
mean(coef(est)[2] > B) mean(coef(est)[2] < abs(B))
```

(2 points)

2:	a other value	b $\beta_x < 0$	c $\beta_x > 1$	d $\beta_x > 0$	e $\beta_x < -1$

Question 3: How could you in the previous question change the last statement (`mean(coef(est)[2] > B)`) to obtain a test for the alternative hypothesis $\beta_x \neq 0$?

(2 points)

3a: None of the following answers are correct.

3b: `mean(B > abs(coef(est)[2]))`

3c: `mean(abs(B) > abs(coef(est)[2]))`

3d: `mean(abs(B) < abs(coef(est)[2]))`

3e: `mean(B != coef(est)[2])`

Question 4: An event can have two possible outcomes, 0 or 1. You are interested in the probability p of obtaining a 1. You assume that p follows a Beta distribution. Your prior is that the parameters of the Beta distribution are $\alpha = \beta = 1$. You observe three times a 1 and once a 0. What is your posterior for α and β? (2 points)

4:	a other value	b $\alpha = 1/3, \beta = 1$	c $\alpha = 1/4, \beta = 1/2$	d $\alpha = 4, \beta = 2$	e $\alpha = 3, \beta = 1$

Question 5: Later you observe one more time a 1 and no 0. Given all your observations, what is now your posterior for β?

(2 points)

5:	a other value	b 1/3	c 3	d 1	e 1/2

Question 6: A random variable X follows a normal distribution $X \sim N(\mu, 2)$ with unknown mean μ and known precision $\tau = 2$. Your prior for μ follows a normal distribution $\mu \sim N(10, 5)$ with mean $\mu_0 = 10$ and precision $\tau_0 = 5$. You use JAGS to estimate the following model (α and β are placeholders):

```
for (i in 1:length(x)) { x[i] ~ dnorm(mu,tau) }
mu ~ dnorm(10,β);
α
```

What should you use for α:

(2 points)

6:	a other value	b `tau <- 2`	c `tau <- 5`	d `tau ~ 2`	e `tau ~ 5`

Question 7: What should you use for β:

(2 points)

7:	a other value	b `tau <- 5`	c `tau <- 2`	d 2	e 5

Question 8: There are two possible states of the world: A and B. Your Null hypothesis is that you are in state A. Depending on the true state A or B your test generates outcomes a and b with the following probabilities:

		outcome	
		a	b
state	A	2/3	1/3
	B	1/2	1/2

Successive test outcomes are independent of each other. You use the following protocol: You apply your test once. If the outcome is b, you stop and reject your Null hypothesis. If the outcome is a, you apply the test again. If this time the outcome is b, you stop and reject your Null hypothesis. Otherwise you stop and do not reject. How probable is it that you reject if you are in state A? (3 points)

8:	a other value	b 2/3	c 1	d 1/3	e 5/9

Question 9: Consider the above situation again. Now your prior is that with equal probability you are in state A or B. You apply your test once and you observe the outcome b. What is your posterior probability to be in state A? (2 points)

9:	a other value	b 4/5	c 1/5	d 2/5	e 3/5

Question 10: What must (in the previous question) your prior $\Pr(A)$ be, such that after the test (where you observe b) you find A and B equally probable? (3 points)

10:	a other value	b $2/5$	c $1/2$	d $3/5$	e $4/5$

Question 11: Y is a binary variable which can be either 0 or 1. $\mathcal{L}$ is the logistic distribution function. You estimate the following model:

$$\Pr(Y = 1|X) = \mathcal{L}(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$$

Your estimate is $\beta_0 = 2$, $\beta_1 = -1$ and $\beta_2 = 3$. What are the estimated odds for $\Pr(Y = 1)/\Pr(Y = 0)$? (2 points)

11:	a other value	b $2 - X_1 + 3X_2$	c $e^{X_1+3X_2}$	d $e^{-2+X_1-3X_2}$	e $e^{2-X_1+3X_2}$

Question 12: Y is a binary variable which can be either 0 or 1. Φ is the cumulative distribution function of the normal distribution. ϕ is the density function of the normal distribution with $\phi(x) = d\Phi(x)/dx$. You estimate the following model:

$$\Pr(Y = 1|X) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$$

Your estimate is $\beta_0 = 2$, $\beta_1 = -1$ and $\beta_2 = 3$. What is the marginal effect of X_2 on $P(Y = 1)$? (2 points)

12:	a other value	b $3\phi(3X_2)$	c $3\phi(2 - X_1 + 3X_2)$	d $2 - X_1 + 3X_2$	e 3

Question 13:
You estimate the effect of x1 and x2 on y with the help of a logit model. You obtain the following output from `glm`:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.29	0.81	0.36	0.7211
x1	2.40	1.06	2.26	0.0241
x2	48.94	19.27	2.54	0.0111

What is the marginal effect of x1 on $P(y = 1)$? (3 points)

13a: None of the following answers are correct.

13b: `2.40*dnorm(0.29+2.40*x1+48.94*x2)`

13c: `2.40*exp(0.29+2.40*x1+48.94*x2)`

13d: `2.40*dlogis(0.29+2.40*x1+48.94*x2)`

13e: `2.40*plogis(0.29+2.40*x1+48.94*x2)`

Question 14: You want to measure the effect of a continuous variable x on an outcome y which can be either y=0 or y=1. With **R**, which command allows you to do this? (2 points)

14a: None of the following answers are correct.

14b: `glm(y ~ x)`

14c: `glm(y ~ x,family=binomial(link=probit))`

14d: `glm(y ~ x,family=probit)`

14e: `glm(y ~ x,family=probit(link=binomial))`

Question 15: You use JAGS to estimate the following model:

```
for (i in 1:length(y)) { y[i] ~ dpois(p[i]) ;
    p[i] <- exp(b[1]+b[2]*x[i]) ; }
for(k in 1:2) { b[k] ~ dnorm(0,.0001) }
```

You obtain the following output:

	Lower95	Median	Upper95	Mean	SD	psrf
b[1]	3.97	4.00	4.03	4.00	0.02	1.0318
b[2]	4.99	5.00	5.03	5.00	0.01	1.0323

Which statement about the average of the posterior is correct? (3 points)

15a: None of the following answers are correct.

15b: The marginal effect of x on $E[Y]$ is `5*log(5*x+4)`.

15c: The marginal effect of x on $E[Y]$ is `exp(5*x+4)`.

15d: The marginal effect of x on $E[Y]$ is `5`.

15e: The marginal effect of x on $E[Y]$ is `5*exp(5*x+4)`.

Question 16: Y is a binary variable which can be either 0 or 1. $\mathcal{L}$ is the logistic distribution function. You estimate the following model:

$$\Pr(Y = 1|X) = \mathcal{L}(\beta_0 + \beta_1 X)$$

Your estimate is $\beta_0 = 0$ and $\beta_1 = 5$. What are the estimated odds for $\Pr(Y = 1)/\Pr(Y = 0)$? (2 points)

16:	a other value	b e^{5X}	c $\mathcal{L}(5X)$	d $5e$	e $\frac{e}{5+e}$

Question 17: You use JAGS to estimate a probit model. Consider the following model (α and β are placeholders):

```
for (i in 1:length(y)) { α ; β }
for(k in 1:2) {b[k] ~ dnorm(0,.0001) }
```

Assume that α has the value `y[i] ~ dbern(p[i])`
What should be the value of β:
(3 points)

17a: None of the following answers are correct.

17b: `p[i] ~ pnorm(b[1]+b[2]*x[i],0,1)`

17c: `p[i] <- dnorm(b[1]+b[2]*x[i],0,1)`

17d: `p[i] ~ dnorm(b[1]+b[2]*x[i],0,1)`

17e: `p[i] <- pnorm(b[1]+b[2]*x[i],0,1)`

Question 18: Now y is a count variable. You want to use a Poisson model. What should be the value of α?
(2 points)

18a: None of the following answers are correct.

18b: `y[i] <- exp(dnorm(p[i]))`

18c: `y[i] ~ dnorm(p[i])`

18d: `y[i] ~ dprobit(p[i])`

18e: `y[i] ~ dpois(p[i])`

Question 19: What should in this case be the value of β?
(2 points)

19a: None of the following answers are correct.

19b: `p[i] ~ exp(b[1]+b[2]*x[i])`

19c: `p[i] <- b[1]+b[2]*x[i]`

19d: `p[i] ~ b[1]+b[2]*x[i]`

19e: `p[i] <- exp(b[1]+b[2]*x[i])`

Question 20: You use JAGS to estimate the following model (α and β are placeholders):

```
for (i in 1:length(y)) { α ; β }
for(k in 1:3) { b[k] ~ dnorm(0,.0001) }
theta ~ dgamma(0.01,0.01)
```

Assume that α has the value `y[i] ~ dpois(p[i])`
What type of dependent variable does one typically estimate with this kind of model?

(3 points)

20a: None of the following answers are correct.

20b: A variable which can take only values larger than one.

20c: A count variable.

20d: A variable which can take negative as well as positive values.

20e: A binary variable.

Question 21: Now you want to estimate a negative binomial model where y is a function of an independent variable x. Assume that in the previous specification β has the following value:

```
p[i] <- theta/(theta+mu[i])
mu[i] <- exp(b[0]+b[1]*x1[i]+b[2]*x2[i])
```

What should now be the value of α?

(2 points)

21a: None of the following answers are correct.

21b: `y[i]~dnegbin(p[i],theta)`

21c: `y[i]~dnegbin(b[0]+b[1]*x1[i]+b[2]*x2[i],theta)`

21d: `y~dnegbin(p[i],theta)`

21e: `y<-dnegbin(p[i],theta)`

Question 22: You use JAGS to estimate the following model:

```
for (i in 1:length(y)) { y[i] ~ dpois(p[i]) ;
    p[i] <- exp(b[1]+b[2]*x[i]) }
for(k in 1:2) { b[k] ~ dnorm(0,.0001) }
```

Let $\Pr(\sigma)$ denote the probability that σ is true. Which statement is correct?

(3 points)

22a: None of the following answers are correct.

22b: $\Pr(\texttt{y[i]} = k) = \lambda^k e^{-\lambda}/k!$ with $\lambda = \texttt{b[1]+x[i]*b[2]}$.

22c: $\Pr(\texttt{y[i]} = k) = \lambda^k e^{-\lambda}/k!$ with $k = \texttt{b[1]+x[i]*b[2]}$.

22d: $\Pr(\texttt{y[i]} = k) = \lambda^k e^{-\lambda}/k!$ with $\lambda = e^{\texttt{b[1]+x[i]*b[2]}}$.

22e: $\Pr(\texttt{y[i]} \geq k) = \lambda^k e^{-\lambda}/k!$ with $k = e^{\texttt{b[1]+x[i]*b[2]}}$.

Question 23: You estimate the following multinomial logit model:

$$\eta_a = x'\beta_a + \xi_a$$
$$\eta_b = x'\beta_b + \xi_b$$
$$\eta_c = x'\beta_c + \xi_c$$

The decision maker chooses alternative $k \in \{a, b, c\}$ if $\eta_k \geq \eta_j$ for all $j \in \{a, b, c\}$.
You obtain the following output in **R**:

```
multinom(formula = y~x1+x2-1)
Coefficients:
  x1 x2
b 4 -3
c -1 -1
```

With the same data you repeat the estimation, now using a different outcome as a reference. You obtain the following output (α_{ij} are placeholders):

```
multinom(formula = y~x1+x2-1)
Coefficients:
  x1  x2
a α11 α12
c α31 α32
```

What do you expect for α_{12}? (3 points)

	a	b	c	d	e
23:	other value	-4	2	3	-5

Question 24: What do you expect for α_{31}? (3 points)

	a	b	c	d	e
24:	other value	2	3	-5	-4

Question 25: You estimated a mixed effects model with `lmer` and you have stored the result in the variable `mer`. You have also estimated the same model, though without the random effects, with `lm` and stored the result in the variable `ols`. To find out whether the mixed effects model is justified, you run the following approximate permuatation test:

```
lls <- replicate(2500,{
    y1 <- simulate(ols)[[1]]
    l.ols <- logLik(lm(y1 ~ x))[1]
    l.mer <- logLik(refit(mer,y1))[1]
    2*(l.mer-l.ols)
})
mean(lls < 2*(logLik(mer)-logLik(ols))[1])
```

The last statement returns the value 0.6. Your Null hypothesis is that there is no random effect. What is the p-value? (3 points)

	a	b	c	d	e
25:	other value	.8	.2	.4	.6

Question 26: You estimate the following mixed effects model

$$y_{it} = \beta_1 + \beta_2 x_{it} + \nu_i + \nu'_j + \epsilon_{it}$$

where $i \in \{1, \ldots, I\}$ denotes the individual, $j \in \{1, \ldots, J\}$ denotes a group, and $t \in \{1, \ldots, T\}$ denotes time. ϵ_{it} is the residual, ν_i is the individual specific random effect, ν'_j is the group specific random effect. The vector `j[i]` tells you to which group individual `i` belongs. You use the following model in JAGS (α is a placeholder):

```
for(k in 1:length(y)){
    y[k] ~ dnorm( beta[1] + beta[2]*x[k] + α, tau[1] )
}
for(k in 1:max(i)) { nu[k] ~ dnorm(0,tau[2]) }
for(k in 1:max(j)) { nu2[k] ~ dnorm(0,tau[3]) }
for(k in 1:2) { beta[k] ~ dnorm(0,.0001) }
for(k in 1:3) { tau[k] ~ dgamma(.01,.01) }
```

Here `y` and `x` are two $I \times T$ long vectors denoting the dependent variable y and the independent x. `i` is a vector with length $I \times T$ denoting to which individual i each observation k belongs. `j` is a vector of length I denoting to which group j each individual i belongs.
What should be the value of α ?

(4 points)

26a: None of the following answers are correct.

26b: `nu[k] + nu2[k]`

26c: `nu[i[k]] + nu2[j[i[k]]]`

26d: `nu[j[i[k]]] + nu2[j[i[k]]]`

26e: `nu[i[k]]] + nu2[i[j[k]]]`

Question 27: You want to estimate the following relationship:

$$\texttt{Y} = \beta_0 + \beta_1 \texttt{X1} + \beta_2 \texttt{X2} + \epsilon$$

You fear that X2 and ϵ might be correlated. You decide to use Z1 and Z2 as instruments for X2. How can you estimate this problem with `ivreg`?

(3 points)

27a: None of the following answers are correct.

27b: `ivreg(Y ~ X1 + X2 | Z1 + Z2)`

27c: `ivreg(Y ~ X1 + X2 + Z1 + Z2 | X1 + Z1 + Z2)`

27d: `ivreg(Y ~ X2 | X1 + Z1 + Z2)`

27e: `ivreg(Y ~ X1 + X2 | X1 + Z1 + Z2)`

Question 28: You want to estimate an equivalent of the following instrumental variables model

```
ivreg(y ~ x1 | x2)
```

Consider the following model for JAGS (α and β are a placeholders):

```
for(i in 1:length(y)) {
  x1[i] ~ dnorm( x1H[i], tau[2])
  x1H[i] <- α
  y[i] ~ dnorm(β, tau[1])
}
for(k in 1:2) {
  b[k] ~ dnorm(0,.0001)
  c[k] ~ dnorm(0,.0001)
  tau[k] ~ dgamma(0.01,.01)
}
```

What is the value of β? (3 points)

28a: None of the following answers are correct.

28b: `x1[i]`

28c: `x1H[i]`

28d: `b[1]+b[2]*x1H[i]`

28e: `b[1]+b[2]*x1[i]`

Question 29: What is the value of α? (3 points)

29a: None of the following answers are correct.

29b: `c[1]*x2`

29c: `c[1]+c[2]*x2[i]`

29d: `c[1]+c[2]*x[2]`

29e: `x2+x3`

Question 30: You want to explain Y as a linear function of either X_1 or X_2. You are unsure which model to use, either $Y = \beta_0 + \beta_1 X_1$ or $Y = \beta_0 + \beta_1 X_2$ You use the following model for model selection in JAGS (α is a placeholder):

```
for (i in 1:length(y)) {
  y[i] ~ dnorm(ifelse(equals(mI,1),α),tau[mI+1])
}
for (j in 1:4) { b[j] ~ dnorm(0,.001) }
for (j in 1:2) { tau[j] ~ dexp(0.01) }
mI ~ dbern(p)
p ~ dunif(0,1)
```

y contains your Y, x1 and x2 contain your X_1 and X_2, respectively. What should you fill in for α?

(3 points)

30a: None of the following answers are correct.

30b: `b[1]+b[2]*x1[i], b[3]+b[4]*x2[i]`

30c: `b[1]+b[2]*x1[i], b[1]+b[2]*x2[i]`

30d: `b[1]+b[2]*x1[i]+b[3]*x[2]`

30e: `y <- b[1]+b[2]*x1[i]`

Question 31:

In your output from the previous model you obtain a mean for `mI` of 0. What are the odds of the model $Y = \beta_0 + \beta_1 X_1$?

(2 points)

	a	b	c	d	e
31:	other value	1/2	0	$e/(1+e)$	1

Question 32:

In your output from JAGS you obtain a 95%-credible interval for `mI` of $CI_{95\%} = [0, 1]$ and a mean for `mI` of 0.2. What are the odds of the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$?

(3 points)

	a	b	c	d	e
32:	other value	1/4	1/5	5	4

3 MW24.1, February 2019 Final

Question 1: Based on N observations of Y, x1 and x2, respectively, you estimate the model $\texttt{Y} = \beta_0 + \beta_1 \texttt{x1} + \beta_2 \texttt{x2} + \epsilon$. You find parameter estimates $\hat\beta_0, \hat\beta_1, \hat\beta_2$ as well as estimated residuals $\hat\epsilon$. You use the bootstrap to estimate the standard deviation of the ratio $\hat\beta_1/\hat\beta_2$. You use the following code:

```
est <- lm(Y ~ x1 + x2)
beta1beta2 <- replicate(1000,{
  estB <- lm(predict(est) + α ~ x1 + x2)
  coef(estB)["x1"]/coef(estB)["x2"]
})
sd(beta1beta2)
```

What value should you use for α? (3 points)

1a: None of the following is correct.

1b: `residuals(sample(est))`

1c: `sample(residuals(est))`

1d: `predict(residuals(est))`

1e: `residuals(predict(est))`

(Remember that `predict` provides $\hat Y = \hat\beta_0 + \hat\beta_1 \texttt{X}_1 + \hat\beta_2 \texttt{X}_2$, `coef` provides $\hat\beta$, and `residuals` provides $\hat\epsilon$.)

Question 2: What value should α have for a parametric bootstrap? (3 points)

2a: None of the following is correct.

2b: `rnorm(N,0,sd(residuals(est)))`

2c: `residuals(N,0,rnorm(sd(est)))`

2d: `sample(sd(residuals(est)))`

2e: `rnorm(sd(residuals(est)))`

Question 3: A random variable follows a normal distribution with unknown precision τ and known mean $\mu = 5$. According to your prior $\tau \sim \Gamma(2,1)$. In your sample x with 8 observations you find `mean(x)=3` and `var(x)=2`. If your posterior $\tau \sim \Gamma(\alpha_{post}, \beta_{post})$, what is then the value of α_{post}? (2 points)

	a	b	c	d	e
3:	other value	4	6	8	3

Question 4: What is the value of β_{post}? (2 points)

	a	b	c	d	e
4:	other value	8	9	1	3

Question 5: Consider the same problem as in the previous two questions. Now you use JAGS to determine τ_{post}. You estimate the following model (γ and δ are placeholders):

```
for (i in 1:length(x)) { x[i] ~ γ }
tau ~ δ
```

What should you use for γ: (2 points)

5a: None of the following is correct.

5b: `dnorm(alpha,beta)`

5c: `dgamma(1,1)`

5d: `dgamma(alpha,beta)`

5e: `dnorm(5,tau)`

Question 6: What should you use for δ: (2 points)

6a: None of the following is correct.

6b: `dnorm(5,tau)`

6c: `dnorm(alpha,beta)`

6d: `dgamma(alpha,beta)`

6e: `dgamma(2,1)`

Question 7: You use JAGS to estimate the following model:

```
for (i in 1:length(x)) { x[i] ~ dnorm(beta,tau) }
beta ~ dnorm (0,.0001);
tau ~ dgamma(.01,.01)
```

You obtain the following output:

```
        Lower95 Median Upper95 Mean  SD  psrf
beta     9.2     11     12.8   11.2 2.2 1.0002
```

Which statements are correct?

(more than one answer possible, 10 points)

7a: None of the following is correct.

7b: With probability 95% the posterior x is between 9.2 and 12.8.

7c: With probability 95% the true mean of x is between 9.2 and 12.8.

7d: The posterior distribution of the true mean of x has a standard deviation of 1.0002

7e: The p-value of this test is 1.0002.

Question 8: A random variable X follows a normal distribution $X \sim N(\mu, 1)$ with unknown mean μ and known precision $\tau = 1$. Your prior for μ follows a normal distribution $\mu \sim N(-3, 1)$ with mean $\mu_0 = -3$ and precision $\tau_0 = 1$. You use JAGS to estimate the following model (α, β and γ are placeholders):

```
for (k in 1:length(x)) { x[k] ~ dnorm(mu,tau) }
mu ~ dnorm (α,β);
γ
```

What should you use for α: (2 points)

	a	b	c	d	e
8:	other value	1	2	-3	-2

Question 9: What should you use for β: (2 points)

	a	b	c	d	e
9:	other value	2	-3	-2	1

Question 10: What should you use for γ: (2 points)

	a	b	c	d	e
10:	other value	`tau <- 3`	`tau <- 1`	`tau ~ 3`	`tau ~ 1`

Question 11: There are two possible states of the world: X and Y. According to your Null hypothesis your are in state X. Depending on the true state of the world your test generates outcomes x and y with the following probabilities:

		outcome	
		x	y
state	X	4/5	1/5
	Y	1/2	1/2

Successive outcomes of this test are independent of each other. You use the following protocol: You apply this test once. If the outcome is y, you stop and reject your Null hypothesis. If the outcome is x, you apply the test a second time. If this time the outcome is y, you stop and reject your Null hypothesis. Otherwise you stop and do not reject. How probable is it that you reject if you are in state X? (3 points)

	a	b	c	d	e
11:	other value	1/5	9/25	5/6	1/25

Question 12: Consider the above situation again. Now your prior is that with equal probability you are in state X or Y. Assume that you apply your test exactly once. You observe the outcome y. What is your posterior probability to be in state Y? (2 points)

	a	b	c	d	e
12:	other value	9/14	5/7	1/10	1/6

Question 13: Assume that you apply the above test exactly once. You observe y. What must your prior $\Pr(X)$ be such that your posterior probabilities for X and Y are the same? (3 points)

	a	b	c	d	e
13:	other value	5/7	1/5	1/2	4/5

Question 14: Y is a binary variable which can be either 0 or 1. $\mathcal{L}$ is the logistic distribution function. You estimate the following model:

$$\Pr(Y = 1|X) = \mathcal{L}(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$$

Your estimate is $\beta_0 = 0$, $\beta_1 = 1$ and $\beta_2 = 2$. What are the estimated odds for $\Pr(Y = 1)/\Pr(Y = 0)$? (2 points)

	a	b	c	d	e
14:	other value	$e^{X_1+2X_2}$	e	$e^{X_1+2X_2}/e^{X_2}$	$\log(X_1+2X_2)$

Question 15: Y is a binary variable which can be either 1 or 0. Φ is the cumulative distribution function of the normal distribution. ϕ is the density function of the normal distribution with $\phi(x) = d\Phi(x)/dx$. You estimate the following model:

$$\Pr(Y = 1|X) = \Phi(\beta_0 + \beta_1 X)$$

Your estimate is $\beta_0 = 2$ and $\beta_1 = 0$. What is the marginal effect of X on $P(Y = 1)$? (2 points)

	a	b	c	d	e
15:	other value	$\Phi(2)$	0	2	$2\phi(2)$

Question 16: You use JAGS to estimate a probit model. Your dependent variable is y, your explanatory variable is x. Consider the following model (α and β are placeholders):

```
for (i in 1:length(y)) { α }
for(k in 1:2) { b[k] ~ dnorm (0,.0001) }
```

What should be the value of α: (3 points)

16a: None of the following is correct.

16b: `y[i] <- pnorm(b[1]+b[2]*x[i],0,1)`

16c: `y[i] ~ dbern(pnorm(b[1]+b[2]*x[i],0,1))`

16d: `y[i] ~ dbern(plogis(b[1]+b[2]*x[i],0,1))`

16e: `y[i] <- dbern(plogis(b[1]+b[2]*x[i],0,1))`

Question 17: Now y is a count variable. You want to use a negative binomial model. You start from the following specification (α is a placeholder):

```
for(i in 1:length(y)) {
  y[i] ~ dnegbin(p[i],theta)
  p <- theta/(theta+mu[i])
  α
}
theta ~ dgamma(0.01,0.01)
for(k in 1:2) { b[k] ~ dnorm(0,.0001) }
```

What should be the value of α? (2 points)

17a: None of the following is correct.

17b: `mu[i] ~ dpois(b[1] + b[2]*x[i])`

17c: `mu[i] ~ exp(b[1] + b[2]*x[i])`

17d: `mu[i] <- exp(b[1] + b[2]*x[i])`

17e: `mu[i] <- dpois(b[1] + b[2]*x[i])`

Question 18: Which statements are correct?
(more than one answer possible, 2 points)

18a: The negative binomial models can be used when the dependent variable takes negative values.

18b: In the previous model the dispersion parameter `theta` is assumed to be the same for all observations.

18c: In the previous model the prior for the dispersion parameter `theta` follows a negative binomial distribution.

18d: When the dispersion parameter `theta` is very large, the negative binomial model approximates the Poisson model.

18e: Only when `theta` is allowed to be different for all observations, the negative binomial model allows us to model a process with a larger number of zeroes than the Poisson model.

Question 19: You estimate the following multinomial logit model:

$$\eta_1 = x'\beta_1 + \xi_1$$
$$\eta_2 = x'\beta_2 + \xi_2$$
$$\eta_3 = x'\beta_3 + \xi_3$$

The decision maker chooses alternative k if $\eta_k \geq \eta_j$ for all j. You expect the following:

$$(\beta_1, \beta_2, \beta_3) = \begin{pmatrix} 0 & 4 & 2 \\ 0 & 0 & 2 \end{pmatrix}$$

You obtain the following output in R (α_1 and α_2 are placeholders):

```
multinom(formula = y ~ x1 + x2 - 1)
Coefficients:
  x1 x2
2  1  0
3 α1 α2
```

What do you expect for α_1? (3 points)

	a	b	c	d	e
19:	other value	0	1/2	2	-1/2

Question 20: What do you expect for α_2? (3 points)

	a	b	c	d	e
20:	other value	1/2	2	-1/2	0

Question 21: You use `lmer` from the `lme4` library to estimate the following mixed effects model:

```
lmer(y ~ x + (x - 1 | i))
```

You obtain the following output from `lmer`:

```
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ x + (x - 1 | i)
Random effects:
Groups   Name Variance Std.Dev. Corr
i        x    0.2946   0.5428
Residual      1.1521   1.0733
Number of obs: 100, groups: i, 10
Fixed effects:
            Estimate Std. Error t value
(Intercept)   3.0276    0.1120   27.03
x            -2.1058    0.2092  -10.07
```

Assume a level of significance of 5%. Based on this output, which of the following statements can you confirm?
(more than one answer possible, 5 points)

21a: None of the following is correct.

21b: Since it is hard to determine degrees of freedom for the t-statistic, tests for fixed effects should be based on a bootstrap.

21c: The model includes a random effect for the slope

21d: The model includes a random effect for the intercept

21e: The model includes a fixed effect for the intercept

Question 22: You estimate the following mixed effects model

$$y_{i,t} = \beta_1 + (\beta_2 + \nu_i)x_{i,t} + \epsilon_{i,t}$$

where $i \in \{1, \ldots, I\}$ denotes the individual, and $t \in \{1, \ldots, T\}$ denotes time. $\epsilon_{i,t}$ is the residual, ν_i is an individual specific random effect. You use the following model in JAGS (α is a placeholder):

```
for(k in 1:length(y)){ α }
for(j in 1:max(i)){nu[j] ~ dnorm(0,tau[2])}
for(k in 1:2){beta[k] ~ dnorm(0,.0001);
  tau[k] ~ dgamma(.01,.01)}
```

Here y and x are two $I \times T$ long vectors denoting the dependent variable y and the independent x. i is a $I \times T$ long vector which denotes the individual i. What should be the value of α?

(3 points)

22a: None of the following is correct.

22b: `y[k]~dnorm(beta[1]+beta[2]*x[k]+nu[i[k]],tau[1])`

22c: `y[k]~dnorm(beta[1]+(beta[2]+nu[i[k]])*x[k],tau[1])`

22d: `y[k]~dnorm(beta[1]+beta[2]*x[k]+nu[i],tau[1])`

22e: `y[k]~dnorm(beta[1]+(beta[2]+nu[i])*x[k],tau[1])`

Question 23: You estimate the following relationship:

$$\mathtt{Y} = \beta_0 + \beta_1 \mathtt{X1} + \beta_2 \mathtt{X2} + \epsilon$$

You fear that X2 and ϵ are correlated. You decide to use X1, Z1 and Z2 as instruments for X2. How can you estimate this model with `ivreg`? (3 points)

23a: None of the following is correct.

23b: `ivreg(Y ~ X1 + X2 | X1 + Z1 + Z2)`

23c: `ivreg(Y ~ X1 + X2 + Z1 + Z2|Z1 + Z2)`

23d: `ivreg(Y ~ X2 | X1 + Z1 + Z2)`

23e: `ivreg(Y ~ X1 + X2 + Z1 + Z2)`

Question 24: The variables x1, x2, and x3 are independent and from a continuous distribution. Which of the following instrumental variable models are underidentified?

(more than one answer possible, 5 points)

24a: `ivreg(y ~ x1 | x1)`

24b: `ivreg(y ~ x1 + x3 | x2 + I(x2^2))`

24c: `ivreg(y ~ x1 | x2 + x3)`

24d: `ivreg(y ~ x1 | x2)`

24e: `ivreg(y ~ x1 + x3 | x2)`

Question 25: You want to estimate an equivalent of the following instrumental variables model

```
ivreg(y ~ x1 + x2 | x2 + x3)
```

Consider the following model for JAGS (α and β are a placeholders):

```
for(i in 1:length(y)) {
  x1[i] ~ dnorm( α, tau[2])
  x1H[i] <- β
  y[i] ~ dnorm(γ, tau[1])
}
for(k in 1:3) {
  b[k] ~ dnorm(0,.0001)
  c[k] ~ dnorm(0,.0001)
  tau[k] ~ dgamma(0.01,.01)
}
```

What is the value of α? (3 points)

25a: None of the following is correct.

25b: `x1H[i] + x2[i]`

25c: `x2H[i]`

25d: `x1H[i]`

25e: `x1H[i] + x2H[i]`

Question 26: What is the value of β? (2 points)

26a: None of the following is correct.

26b: `c[1] + c[2] * x2[i] + c[3] * x3[i]`

26c: `c[1] + c[2] * x3[i]`

26d: `c[1] + c[2] * x2H[i] + c[3] * x3[i]`

26e: `dnorm(c[1] + c[2] * x3[i],tau[1])`

Question 27: What is the value of γ? (2 points)

27a: None of the following is correct.

27b: `b[1] + b[2] * x1H[i]`

27c: `b[1] + b[2] * x1H[i] + b[3] * x2[i]`

27d: `b[1] + b[2] * x1H[i] + b[3] * x2H[i]`

27e: `b[1] + b[2] * x1H[i] + x2[i]`

Question 28: You estimate the following equation:

$$Y = \sum_{i=0}^{k} \beta_i X^i + \epsilon$$

The following table shows the AIC of the estimated model for different values of k:

k	1	2	3	4	5	6	7
AIC	1.67	-7.78	-11.10	-10.98	-9.10	-8.75	-8.72

Which value of k should you choose, based on the AIC?

(2 points)

28:	a other value	b 3	c 6	d 1	e 2

Question 29: You use the following model for model selection in JAGS (α is a placeholder):

```
for (i in 1:length(y)) {
    y[i] ~ dnorm( α (equals(mI,0),
        b[1] + b[2] * x1[i],
        b[3] + b[4] * x2[i]),
      tau[mI+1])
}
for (j in 1:4) { b[j] ~ dnorm(0,.001) }
for (j in 1:2) { tau[j] ~ dexp(0.01)}
mI ~ dbern(p)
p ~ dunif(0,1)
```

What should you fill in for α? (3 points)

29:	a other value	b exp	c ifelse	d dpois	e p

Question 30: Which econometric models does this JAGS model compare?

(2 points)

30a: None of the following is correct.

30b: $Y = \beta_0 + \beta_1 X_1 + \epsilon$ against $Y = \beta_0 + \beta_1 X_2 + \epsilon$

30c: $Y = \beta_0 + \beta_1 X_1 + \epsilon$ against $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$

30d: $Y = \beta_0 + \beta_1 X_1 + \epsilon$ against $Y = \beta_0 + \beta_1 X_1 - \beta_2 X_2 + \epsilon$

30e: $Y = \beta_0 + \beta_1 X_1 + \epsilon$ against $Y = \beta_0 + \epsilon$

Question 31: In your output from JAGS you obtain a 95%-credible interval for `mI` of $CI_{95\%} = [0, 1]$ and a mean for `mI` of 0.8. What can you conclude? (3 points)

31a: None of the following is correct.

31b: The probability of the model $Y = \beta_0 + \beta_1 X_1 + \epsilon$ is 0.2.

31c: The probability of the model $Y = \beta_0 + \beta_1 X_1 + \epsilon$ is 1.

31d: Since the credible interval for `mI` includes the extremes 0 and 1 one can not make a statement how probable the models are.

31e: The probability of the model $Y = \beta_0 + \beta_1 X_1 + \epsilon$ is 0.8.

---

4 MW24.1, April 2018 Final Resit

Question 1: You have estimated the model $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ and found parameter estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ as well as estimated residuals $\hat{\epsilon}_i$.
You want to use a bootstrap to estimate the interquartile range (the difference between the 75th and the 25th percentile) of $\hat{\beta}_1$. To do this, you draw a large number B of bootstrap replications $\hat{\beta}_1^{*k}$ and use their interquartile range $\text{IQR}_{k=1}^{B}(\hat{\beta}_1^{*k})$ as your estimator. You obtain these $\hat{\beta}_1^{*k}$ from estimating the following equation:

$$Y_i^{*k} = \beta_0^{*k} + \beta_1^{*k} X_i + \epsilon_i^k$$

What should you use for Y_i^{*k}? (in the following the function "`sample`(x)" is a function that returns a sample of x in a random order but with the same size as x.) (3 points)

1a: None of the following answers are correct.

1b: $Y_i^{*k} = \texttt{sample}(\hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\epsilon}_i)$

1c: $Y_i^{*k} = \hat{\beta}_0 + \hat{\beta}_1 X_i + \texttt{sample}(\hat{\epsilon}_i)$

1d: $Y_i^{*k} = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\epsilon}_i$

1e: $Y_i^{*k} = \hat{\beta}_0 + \hat{\beta}_1 \texttt{sample}(X_i) + \hat{\epsilon}_i$

Question 2:
You have estimated the following Poisson model with R.

```
est <- glm(y ~ x1 + x2,family=poisson(link=log))
```

You want to use a bootstrap to estimate the standard deviation of the coefficient of `x1`. To do this, you draw a large number B of bootstrap replications $\hat{\beta}_1^{*k}$ and use their standard deviation $\text{sd}_{k=1}^{B}(\hat{\beta}_1^{*k})$ as your estimator. You obtain these $\hat{\beta}_1^{*k}$ as follows:

```
N<-length(y)
betaHat1<- replicate(B, Yh<-rpois(N,α); coef(glm(β))[2] )
```

What is the value of α? (3 points)

2a: None of the following answers are correct.

2b: `exp(predict(est))`

2c: `log(predict(est))`

2d: `exp(coef(est))`

2e: `log(coef(est))`

Question 3: What is the value of β? (2 points)

3a: None of the following answers are correct.

3b: `Yh ~ x1+x2,family=poisson(link=log)`

3c: `Yh ~ x1+x2,family=binomial(link=poisson)`

3d: `Yh ~ x1+x2,family=binomial(link=logit)`

3e: `Yh ~ x1+x2,family=logit(link=poisson)`

Question 4: A random variable follows a normal distribution with unknown precision τ and known mean $\mu = 5$. According to your prior $\tau \sim \Gamma(1, 1)$. In your sample `x` with 4 observations you find `mean(x)=3` and `var(x)=4`. If your posterior $\tau \sim \Gamma(\alpha_{post}, \beta_{post})$, what is then the value of α_{post}? (2 points)

4:	a	b	c	d	e
	other value	1	3	9	1/4

Question 5: What is the value of β_{post}? (2 points)

5:	a	b	c	d	e
	other value	8	9	1	3

Question 6: Consider the same problem as in the previous two questions. Now you use JAGS to determine τ_{post}. You estimate the following model (γ and δ are placeholders):

```
for (i in 1:length(x)) { x[i] ~ γ }
tau ~ δ
```

What should you use for γ: (2 points)

6a: None of the following answers are correct.

6b: `dnorm(alpha,beta)`

6c: `dgamma(1,1)`

6d: `dgamma(alpha,beta)`

6e: `dnorm(5,tau)`

Question 7: What should you use for δ: (2 points)

7a: None of the following answers are correct.

7b: `dgamma(1,1)`

7c: `dgamma(alpha,beta)`

7d: `dnorm(5,tau)`

7e: `dnorm(alpha,beta)`

Question 8: With your data `x` you use JAGS to estimate the following model:

```
for (i in 1:length(x)) { x[i] ~ dnorm(mu,tau) }
tau ~ dgamma(1,1)
mu ~ dnorm(0,1)
```

You obtain the following results:

	Lower95	Median	Upper95	Mean	SD	psrf
mu	2.64	3.90	4.95	3.85	0.60	1.00
tau	0.07	0.20	0.38	0.21	0.08	1.00

Which statements are correct?
(more than one answer possible, 10 points)

8a: The p-value of this test is almost 100%

8b: With probability 95% the posterior `x` is between 2.64 and 4.95.

8c: With probability 95% the posterior population mean of `x` is between 2.64 and 4.95.

8d: The posterior probability that the variance of `x` is larger than 0.2 is 1/2.

8e: `psrf` is very close to 1 which suggests that the samplers did converge.

Question 9: Y is a binary variable which can be either 0 or 1. $\mathcal{L}$ is the logistic distribution function. You estimate the following model:

$$\Pr(Y = 1|X) = \mathcal{L}(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$$

Your estimate is $\beta_0 = 3$, $\beta_1 = 1$ and $\beta_2 = -2$. What are the estimated odds for $\Pr(Y = 1)/\Pr(Y = 0)$? (2 points)

	a	b	c	d	e
9:	other value	$e^{X_1-2X_2}$	e^{X_1}/e^{2X_2}	$e^{3+X_1-2X_2}$	$3+X_1-2X_2$

Question 10: Y is a binary variable which can be either 0 or 1. Φ is the cumulative distribution function of the normal distribution. ϕ is the density function of the normal distribution with $\phi(x) = d\Phi(x)/dx$. You estimate the following model:

$$\Pr(Y = 1|X) = \Phi(\beta_0 + \beta_1 X_1 + \beta_2 X_2)$$

Your estimate is $\beta_0 = 3$, $\beta_1 = 1$ and $\beta_2 = -2$. What is the marginal effect of X_2 on $P(Y = 1)$? (2 points)

	a	b	c	d	e
10:	other value	$-2\phi(3+X_1-2X_2)$	$3+X_1-2X_2$	-2	$-2\phi(-2X_2)$

Question 11: You estimate the effect of `x1` and `x2` on `y` with the help of a logit model. You obtain the following output from `glm`:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.21	0.64	-0.33	0.7401
x1	1.37	0.92	1.49	0.1367
x2	28.15	9.80	2.87	0.0041

What is the marginal effect of `x1` on $P(y = 1)$? (3 points)

11a: None of the following answers are correct.

11b: `1.37*dnorm(-0.21+1.37*x1+28.15*x2)`

11c: `1.37*exp(-0.21+1.37*x1+28.15*x2)`

11d: `1.37*dlogis(-0.21+1.37*x1+28.15*x2)`

11e: `1.37*plogis(-0.21+1.37*x1+28.15*x2)`

Question 12: You want to measure the effect of a continuous variable x on an outcome y which can be either y=0 or y=1. With **R**, which command allows you to do this? (2 points)

12a: None of the following answers are correct.

12b: `glm(y ~ x,family=binomial(link=probit))`

12c: `glm(y ~ x,family=probit)`

12d: `glm(y ~ x,family=probit(link=binomial))`

12e: `glm(y ~ x)`

Question 13: You use JAGS to estimate the following model:

```
for (i in 1:length(y)) { y[i] ~ dbern(p[i]) ;
    p[i] <- plogis(b[1]+b[2]*x[i],0,1) ; }
for(k in 1:2) { b[k] ~ dnorm (0,.0001) }
```

You obtain the following output:

```
     Lower95  Median Upper95    Mean
b[1] -6.0000 -5.0000 -4.0000 -5.0000
b[2]  3.0000  5.0000  7.0000  5.0000
```

Which statement about the average of the posterior is correct? (3 points)

13a: None of the following answers are correct.

13b: The odds of $y = 1$ is `exp(5*x-5)`.

13c: The odds of $y = 1$ is `exp(5)`.

13d: The marginal effect of x on $P(y = 1)$ is `5*dnorm(5*x-5)`.

13e: The marginal effect of x on $P(y = 1)$ is `dnorm(5*x-5)`.

Question 14: You use JAGS to estimate the following model (α and β are placeholders):

```
for (i in 1:length(y)) { α ; β }
for(k in 1:3) {b[k]~dnorm (0,.0001) }
theta ~ dgamma(0.01,0.01)
```

Assume that α has the value `y[i]~dpois(p[i])`

What type of dependent variable does one typically estimate with this kind of model? (3 points)

14a: None of the following answers are correct.

14b: A count variable.

14c: A variable which can take negative as well as positive values.

14d: A binary variable.

14e: A variable which can take only values larger than one.

Question 15: Now you want to estimate a negative binomial model where y is a function of an independent variable x. Assume that in the previous specification β has the following value:

```
p[i] <- theta/(theta+mu[i])
mu[i] <- exp(b[0]+b[1]*x1[i]+b[2]*x2[i])
```

What should now be the value of α? (2 points)

15a: None of the following answers are correct.

15b: `y[i]~dnegbin(b[0]+b[1]*x1[i]+b[2]*x2[i],theta)`

15c: `y~dnegbin(p[i],theta)`

15d: `y<-dnegbin(p[i],theta)`

15e: `y[i]~dnegbin(p[i],theta)`

Question 16: You estimate the following multinomial logit model:

$$\eta_1 = x'\beta_1 + \xi_1$$
$$\eta_2 = x'\beta_2 + \xi_2$$
$$\eta_3 = x'\beta_3 + \xi_3$$

The decision maker chooses alternative k if $\eta_k \geq \eta_j$ for all j. You do not know the variance of ξ_1, ξ_2, ξ_3. You expect the following:

$$(\beta_1, \beta_2, \beta_3) = \begin{pmatrix} 6 & 2 & -1 \\ 0 & 0 & 1 \end{pmatrix}$$

You obtain the following output in R (α_1 and α_2 are placeholders):

```
multinom(formula = y ~ x1+x2-1)
Coefficients:
    x1   x2
2  -12    0
3   α1   α2
```

What do you expect for α_1? (3 points)

16:	a other value	b 0	c 3	d 6	e -21

Question 17: What do you expect for α_2? (3 points)

17:	a other value	b 3	c 6	d -21	e 0

Question 18:
You use `lmer` from the `lme4` library to estimate two mixed effects models, `m1.mer` and `m2.mer`.

```
m1.mer<-lmer(y~x+(1|i))
m2.mer<-lmer(y~x+(x|i))
```

A comparison of the two models with the help of `anova` yields the following result:

	Df	AIC	BIC	logLik	deviance	χ^2	χ^2 Df	Pr(> χ^2)
m1.mer	4	298	309	-145	290			
m2.mer	6	291	307	-139	279	11.23	2	0.0036

Assume a level of significance of 5%. Which of the following statements is true?

(more than one answer possible, 10 points)

18a: The random effect for the slope is significant.

18b: `anova` does a Laplace test here.

18c: The method used by `anova` here is often too conservative. The true p-value may well be smaller than 0.0036.

18d: The method used by `anova` here is often anti-conservative. The true p-value may well be larger than 0.0036.

18e: It is possible to obtain a better estimate for the p-value of this test with the help of a bootstrap.

Question 19: You obtain the following output from `lmer`:

```
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ x + (x | i)
Random effects:
Groups   Name        Variance Std.Dev.  Corr
i        (Intercept)  0.05504  0.2346
         x            0.30216  0.5497 -1.00
Residual              0.85979  0.9272
Number of obs: 100, groups: i, 10
Fixed effects:
            Estimate Std. Error t value
(Intercept)   2.7642     0.1283  21.544
x            -2.0805     0.2139  -9.728
```

Which of the following statements is true?

(more than one answer possible, 5 points)

19a: None of the following answers are correct.

19b: The random effect for the slope is significant.

19c: The model includes a random effect for the slope

19d: The model includes a random effect for the intercept

19e: The model includes a fixed effect for the intercept

Question 20: You estimate the following mixed effects model

$$y_{i,t} = \beta_1 + \beta_2 x_{i,t} + \nu_i x_{i,t} + \epsilon_{i,t}$$

where $i \in \{1,\ldots,I\}$ denotes the individual, and $t \in \{1,\ldots,T\}$ denotes time. For each individual you have T many observations. $\epsilon_{i,t}$ is the residual, ν_i is the individual specific random effect. You use the following model in JAGS (α is a placeholder):

```
for(it in 1:length(y)){
    y[it] ~ dnorm(beta[1]+beta[2]*x[it]+ α,tau[2] )
}
for(β){nu[j]~dnorm(0,tau[1])}
for(k in 1:2){beta[k]~dnorm(0,.0001);tau[k]~dgamma(.01,.01)}
```

Here `y` and `x` are two $I \times T$ long vectors denoting the dependent variable y and the independent x. `i` is a $I \times T$ long vector denoting the individual i. α and β are placeholders.
What should be the value of α?

(3 points)

20a: None of the following answers are correct.

20b: `nu[it[i]]`

20c: `x[it]*nu[i[it]]`

20d: `x[it]*nu[i[it]]+nu[i[it]]`

20e: `nu[i[it]]`

Question 21: What should be the value of β?

(2 points)

21a: None of the following answers are correct.

21b: `j in 1:length(i)`

21c: `j in 1:length(it)`

21d: `j in 1:max(i)`

21e: `j in 1:max(it)`

Question 22: You want to estimate the equation $Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \epsilon$. You fear that `X2` and ϵ may be correlated. You decide to use `Z` as an instrument for `X2`. You estimate the following regression:

`ivreg(Y ~ X1 + X2 | α)`

What should be the value of α?

(2 points)

22a: None of the following answers are correct.

22b: `X1 + X2 + Z`

22c: `Z`

22d: `X1`

22e: `X1 + Z`

Question 23: You still want to estimate the model

$$\mathtt{Y} = \beta_0 + \beta_1 \mathtt{X1} + \beta_2 \mathtt{X2} + \epsilon$$

still using Z as an instrument for X2. Consider the following model for JAGS (α and β are placeholders):

```
for(i in 1:length(y)) {
  x2[i] ~ dnorm(xHat[i], tau[2])
  xHat[i] α
  y[i] β
}
for(k in 1:4) {
  b[k] ~ dnorm(0,.0001)
  g[k] ~ dnorm(0,.0001)
  tau[k] ~ dexp(0.01)
}
```

What is the value of α? (3 points)

23a: None of the following answers are correct.

23b: `<- g[1]+g[2]*z[i]+g[3]*x2[i]`

23c: `~ g[1]+g[2]*z[i]+g[3]*x2[i]`

23d: `<- z[i]`

23e: `~ z[i]`

Question 24: What is the value of β? (3 points)

24a: None of the following answers are correct.

24b: `<- dnorm(b[1]+b[2]*x1[i]+b[3]*xHat[i],tau[1])`

24c: `~ dnorm(b[1]+b[2]*x1[i]+b[3]*z[i],tau[1])`

24d: `<- dnorm(b[1]+b[2]*x1[i]+b[3]*z[i],tau[1])`

24e: `~ dnorm(b[1]+b[2]*x1[i]+b[3]*xHat[i],tau[1])`

---

5 MW24.1, February 2018 Final

Question 1: To assess income inequality you consider the ratio of the median income to the 10% quantile. The following function determines, hence, inequality in your sample:

```
ineq <- function(x) median(x)/quantile(x,.1)
```

The variable `x` contains the income, so `ineq(x)` yields a measure for inequality. Which command estimates the standard deviation of this measure: (3 points)

1a: None of the following answers are correct.

1b: `replicate(N,sd(ineq(sample(x,replace=TRUE))))`

1c: `replicate(N,sd(ineq(sample(x,replace=FALSE))))`

1d: `sd(replicate(N,ineq(sample(x,replace=FALSE))))`

1e: `sd(replicate(N,ineq(sample(x,replace=TRUE))))`

Question 2: According to your Null hypothesis the distribution of income follows a Gamma distribution with shape and rate parameters 100 and 10. You can create a sample of n pseudo random numbers of such a distribution with the command `rgamma(n,100,10)`. Your alternative hypothesis is that your sample is drawn from a population with a larger measure of inequality. You use the following command to determine a p-value to test your hypothesis (The variable x contains the income):

```
n <- length(x)
sampleIneq <- ineq(x)
mean(α(10000,ineq(rgamma(n,100,10)) β sampleIneq))
```

(3 points)

	a	b	c	d	e
2:	other value	α: sample β: <	α: replicate β: >	α: replicate β: <	α: sample β: >

Question 3: A random variable follows a normal distribution with known precision $\tau = 1$ and mean μ. According to your prior $\mu \sim N(0, 1)$. In your sample x with 5 observations you find `mean(x)=3` and `var(x)=1/4`. If your posterior $\mu \sim N(\mu_{post}, \tau_{post})$, what is then the value of μ_{post}? (3 points)

	a	b	c	d	e
3:	other value	1/2	60/21	6	5/2

Question 4: What is the value of τ_{post}? (2 points)

	a	b	c	d	e
4:	other value	1	4	6	1/4

Question 5: Consider the same problem as in the previous two questions. Now you use JAGS to determine μ_{post}. You estimate the following model (α and β are placeholders):

```
for (i in 1:length(x)) { x[i] ~ dnorm(mu,α) }
mu~dnorm (0,β);
```

What should you use for α: (2 points)

	a	b	c	d	e
5:	other value	1	4	0	1/4

Question 6: What should you use for β: (2 points)

	a	b	c	d	e
6:	other value	0	1/4	1	4

Question 7: In a linear regression model you explain income of a worker as a function of education and a dummy for each region. You have data about workers which are from n different regions, hence you have n different dummies. For each dummy separately you test the Null hypothesis that the coefficient for this dummy is zero. Your level of significance is α. If the coefficients in the population are, indeed, all zero, how probable is it to find one or more statistically significant dummies in your sample? (3 points)

	a	b	c	d	e
7:	other value	$n \cdot \alpha$	α^n	$1-(1-\alpha)^n$	$1-(1-\alpha^n)$

Question 8: Y is a binary variable which can be either 0 or 1. $\mathcal{L}$ is the logistic distribution function. You estimate the following model:

$$\Pr(Y = 1|X) = \mathcal{L}(\beta_0 + \beta_1 X)$$

Your estimate is $\beta_0 = 0$ and $\beta_1 = 2$. What are the estimated odds for $\Pr(Y = 1)/\Pr(Y = 0)$? (2 points)

	a	b	c	d	e
8:	other value	$\frac{e}{1+e}$	e^{2X}	$\mathcal{L}(2X)$	e

Question 9: Y is a binary variable which can be either 0 or 1. Φ is the cumulative distribution function of the normal distribution. ϕ is the density function of the normal distribution with $\phi(x) = d\Phi(x)/dx$. You estimate the following model:

$$\Pr(Y = 1|X) = \Phi(\beta_0 + \beta_1 X)$$

Your estimate is $\beta_0 = 0$ and $\beta_1 = 2$. What is the marginal effect of X on $P(Y = 1)$? (2 points)

	a	b	c	d	e
9:	other value	2	$2X$	$2\phi(2X)$	$\Phi(2X)$

Question 10: Estimating the effect of x on y with different data but with the same model you obtain the following output from `glm`:

```
             Estimate Std. Error z value Pr(>|z|)
(Intercept)      0.3       0.35    0.91  0.3630
          x     1.21       0.49    2.46  0.0141
```

What is the marginal effect of x on $P(y=1)$? (3 points)

10a: None of the following answers are correct.

10b: `1.21*exp(1.21*x+0.3)`

10c: `1.21*dnorm(1.21*x+0.3)`

10d: `dnorm(1.21*x+0.3)`

10e: `pnorm(1.21*x+0.3)`

Question 11: You want to measure the effect of a continuous variable x on an outcome y which can be either y=0 or y=1. With **R**, which command allows you to do this? (2 points)

11a: None of the following answers are correct.

11b: `glm(y ~ x)`

11c: `glm(y ~ x,family=binomial(link=logit))`

11d: `glm(y ~ x,family=probit)`

11e: `glm(y ~ x,family=probit(link=binomial))`

Question 12: You use JAGS to estimate the following model:

```
for (i in 1:length(y)) { y[i] ~ dbern(p[i]) ;
   p[i] <- pnorm(b[1]+b[2]*x[i],0,1) ; }
for(k in 1:2) { b[k] ~ dnorm (0,.0001) }
```

You obtain the following output:

```
      Lower95  Median Upper95    Mean
b[1] -3.0000 -2.0000 -1.0000 -2.0000
b[2]  1.0000  2.0000  3.0000  2.0000
```

Which statement about the average posterior is correct? (3 points)

12a: None of the following answers are correct.

12b: The marginal effect of x on $P(y=1)$ is `dnorm(2*x-2)`.

12c: The odds of $y=1$ is `exp(2*x-2)`.

12d: The odds of $y=1$ is `exp(2)`.

12e: The marginal effect of x on $P(y=1)$ is `2*dnorm(2*x-2)`.

Question 13: You want to compare two models, a Poisson model and a negative binomial model. Remember that the Poisson regression is a special case of the negative binomial regression where $\theta \to \infty$. To keep the notation simple, we assume here that both models regress x on only a constant. Your Null hypothesis is that the data x can be described by a Poisson model. To test this hypothesis and to obtain a p-value you use the following commands (α and β are placeholders, `glm.nb()$theta` extracts the θ from the negative binomial regression, `rpois(`n`,`λ`)` generates pseudo random numbers following a Poisson distribution).

```
N<-length(x)
λ<-coef(glm(x ~ 1))
θ<-glm.nb(x ~ 1)$theta
mean(replicate(10000, α(rpois(N,λ) ~ 1)$theta β θ))
```

(3 points)

13:

a	b	c	d	e
other value	α: glm, β: >	α: glm, β: <	α: glm.nb, β: >	α: glm.nb, β: <

Question 14: You estimate the following multinomial logit model:

$$\eta_a = x'\beta_a + \xi_a$$
$$\eta_b = x'\beta_b + \xi_b$$
$$\eta_c = x'\beta_c + \xi_c$$

The decision maker chooses alternative $k \in \{a,b,c\}$ if $\eta_k \geq \eta_j$ for all $j \in \{a,b,c\}$.

You obtain the following output in **R**:

```
multinom(formula = y~x1+x2-1)
Coefficients:
  x1 x2
a -5 -1
b -2  1
```

With the same data you repeat the estimation, now using a different outcome as a reference. You obtain the following output (α_{ij} are placeholders):

```
multinom(formula = y~x1+x2-1)
Coefficients:
   x1  x2
a α11 α12
c α31 α32
```

What do you expect for α_{12}? (3 points)

14:

a	b	c	d	e
other value	2	-3	-2	0

Question 15: What do you expect for α_{31}? (3 points)

15:

a	b	c	d	e
other value	-3	-2	0	2

Question 16: Now you use JAGS to estimate a multinomial model. You explain a choice among L alternatives. `X` is the matrix of explanatory variables, β is the matrix of coefficients you want to estimate. `K` is the number of columns of `X` (i.e. the number of explanatory variables). Consider the following model for JAGS (α is a placeholder):

```
for (i in 1:length(y)) {
  for (j in 1:L) { Xb[i,j] <- exp(inprod(β[,j],X[i,]))}
  y[i] ~ dcat(Xb[i,1:L])}
α
for (j in 1:L) {for (k in 1:K) { β[k,j] ~ dnorm(0,.0001)}}
```

If your reference category is the first category, what is the value of α? (2 points)

16a: None of the following answers are correct.

16b: `for (k in 1:K) {β[1,k] <- 0}`

16c: `for (j in 1:L) {β[j,1] <- 0}`

16d: `for (j in 1:L) {β[1,j] <- 0}`

16e: `for (k in 1:K) {β[k,1] <- 0}`

Question 17: Consider the following output from **R**:

```
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ x + (x | g)
Random effects:
Groups   Name        Variance Std.Dev.  Corr
g        (Intercept)  0.17927  0.4234
         x            0.02544  0.1595 -1.00
Residual              0.46131  0.6792
Number of obs: 19, groups: g, 4
Fixed effects:
            Estimate Std. Error t value
 (Intercept) -0.01127    0.27435  -0.041
          x  0.95451    0.19546   4.883
```

(more than one answer possible, 5 points)

17a: None of the following answers are correct.

17b: The model can be written as $y_{it} = \beta_0 + \nu_i + \beta_1 x_{it} + \epsilon_{it}$

17c: The model can be written as
$y_{it} = \beta_0 + \nu_i + (\beta_1 + \nu_i')x_{it} + \epsilon_{it}$

17d: The model includes a random effect only for the slope.

17e: The model includes a random effect both for slope and intercept.

Question 18: You estimate the following model with mixed effects:

$$y_{it} = \beta_1 + \beta_2 x_{it} + \nu_i + \nu_i' x_{it} + \epsilon_{it}$$

For each of `I` many individuals you have `T` many observations, hence you have `I`×`T` many observations. To represent this problem in JAGS you use the following model (α is a placeholder):

```
for(it in 1:length(y)){
  y[it]~dnorm(beta[1]+beta[2]*x[it]+α,tau[3])
}
for(k in 1:3){
  beta[k]~dnorm(0,.0001)
  tau[k]~dgamma(.01,.01)
  for(j in 1:max(i)) {nu[k,j]~dnorm(0,tau[k]) }
}
```

`i` identifies to which individual each observation belongs. What should be the value of α ? (3 points)

18a: None of the following answers are correct.

18b: `nu[1,it[i]]+x[it]*nu[2,it[i]]`

18c: `nu[1,i[it]]+x[it]*nu[2,i[it]]`

18d: `nu[1,i[it]]`

18e: `x[it]*nu[2,i[it]]`

Question 19: You estimate the following relationship:

$$\texttt{Y} = \beta_0 + \beta_1 \texttt{X1} + \beta_2 \texttt{X2} + \epsilon$$

You fear that `X1` and ϵ are correlated. You decide to use `X2`, `Z1` and `Z2` as instruments for `X1`. How can you estimate this model with `ivreg`? (3 points)

19a: None of the following answers are correct.

19b: `ivreg(Y ~ X1+X2 | X2+Z1+Z2)`

19c: `ivreg(Y ~ X1+X2+Z1+Z2)`

19d: `ivreg(Y ~ X1+X2+Z1+Z2|Z1+Z2)`

19e: `ivreg(Y ~ X1+X2 | Z1+Z2)`

Question 20: You still want to estimate the model

$$\texttt{Y} = \beta_0 + \beta_1 \texttt{X1} + \beta_2 \texttt{X2} + \epsilon$$

still using `X2`, `Z1` and `Z2` as instruments for `X1`. Consider the following model for JAGS (α and β are placeholders):

```
for(i in 1:length(y)) {
  x1[i] ~ dnorm(xHat[i], tau[2])
  xHat[i] α
  y[i] β
}
for(k in 1:4) {
  b[k] ~ dnorm(0,.0001)
  g[k] ~ dnorm(0,.0001)
  tau[k] ~ dexp(0.01)
}
```

What is the value of α? (3 points)

20a: None of the following answers are correct.

20b: `~ g[1]+g[2]*z1[i]+g[3]*z2[i]+g[4]*x2`

20c: `<- b[1]+b[2]*x1[i]+b[3]*x2[i]`

20d: `~ b[1]+b[2]*x1[i]+b[3]*x2[i]`

20e: `<- g[1]+g[2]*z1[i]+g[3]*z2[i]+g[4]*x2`

Question 21: What is the value of β? (3 points)

21a: None of the following answers are correct.

21b: `<- g[1]+g[2]*z1[i]+g[3]*z2[i]`

21c: `~ dnorm(g[1]+g[2]*z1[i]+g[3]*z2[i])`

21d: `~ dnorm(b[1]+b[2]*xHat[i]+b[3]*x2[i],tau[1])`

21e: `<- dnorm(b[1]+b[2]*xHat[i]+b[3]*x2[i],tau[1])`

Question 22: You want to explain Y as a function of X_1. You are unsure which model to use: either $Y = \beta_0 + \beta_1 X_1$ or $Y = \beta_0 + \beta_1 \log(X_1)$ You use the following model for model selection in JAGS (α is a placeholder):

```
for (i in 1:length(y)) {
  y[i] ~ dnorm(ifelse(equals(mI,1),α),tau[mI+1])
}
for (j in 1:4) { b[j] ~ dnorm(0,.001) }
for (j in 1:2) { tau[j] ~ dexp(0.01) }
mI ~ dbern(p)
p ~ dunif(0,1)
```

`y` contains your Y, `x1` contain your X_1. What should you fill in for α? (3 points)

22a: None of the following answers are correct.

22b: `y[i] <- b[1]+b[2]*log(x1[i])`

22c: `b[1]+b[2]*x1[i], b[3]+b[4]*log(x1[i])`

22d: `b[1]+b[2]*x1[i], b[1]+b[2]*log(x2[i])`

22e: `b[1]+b[2]*x1[i]+b[3]*log(x1[i])`

Question 23: In your output from the previous model you obtain a mean for `mI` of .4. What are the odds of the model $Y = \beta_0 + \beta_1 X_1$? (2 points)

	a	b	c	d	e
23:	other value	0	1/2	2/3	$e^{0.4}$

6 MW24.1, April 2017 Final Resit

Unless stated otherwise always use a significance level of 5%.

Question 1: You obtain the following output from **R**:

```
Call: glm(formula = y ~ x, family = poisson(link = log))
Coefficients:
 (Intercept)      x
          -1      2
```

What is the marginal effect of x on y?

(3 points)

	a	b	c	d	e
1:	other value	2	e^{2x-1}	$2x-1$	$2e^{2x-1}$

Question 2: You estimate two coefficients, β_0 and β_x, as follows:

```
est<-glm(y~x,family=poisson(link=log))
```

For which alternative hypothesis does the following command approximate a p-value?

```
B<-replicate(10000,{
  est<-glm(y~sample(x),family=poisson(link=log))
  coef(est)[2] })
mean(B>coef(est)[2])
```

(2 points)

	a	b	c	d	e
2:	other value	$\beta_x < 0$	$\beta_0 < 0$	$\beta_x > 0$	$\beta_0 > 0$

Question 3: How would you in the previous question change the last statement (`mean(B>coef(est)[2])`) to obtain a two-sided test?

(2 points)

3a: None of the following answers are correct.

3b: `mean(abs(B) < abs(coef(est)[2]))`

3c: `mean(B != coef(est)[2])`

3d: `mean(B > abs(coef(est)[2]))`

3e: `mean(abs(B) > abs(coef(est)[2]))`

Question 4: You have estimated the following model in **R**:

```
est <- lm(y ~ x)
```

You want to use a bootstrap to estimate the variance of the coefficient of x. To do this, you use the following command (α is a placeholder):

```
var(replicate(10000,coef(lm(α))[2]))
```

What should you use for α? (Remember that `predict(est)` generates the predicted values $\hat{Y} = \beta_0 + \beta_1 X$). (3 points)

4a: None of the following answers are correct.

4b: `predict(est)+sample(residuals(est))~x`

4c: `predict(est)+residuals(est)~x`

4d: `sample(predict(est))+residuals(est)~x`

4e: `predict(est)+residuals(est)~sample(x)`

Question 5: A random variable X follows a normal distribution $X \sim N(\mu, 3)$ with unknown mean μ and known precision $\tau = 3$. Your prior of μ follows a normal distribution $\mu \sim N(10, 2)$ with mean $\mu_0 = 10$ and precision $\tau_0 = 2$. You use JAGS to estimate the following model (α and β are placeholders):

```
for (i in 1:length(x)) { x[i] ~ dnorm(mu,tau) }
mu~dnorm (10,β);
α
```

What should you use for α:

(2 points)

5a: None of the following answers are correct.

5b: `tau ~ 3`

5c: `tau ~ dnorm(mu,tau)`

5d: `tau <- 2`

5e: `tau <- 3`

Question 6: A random variable X follows a normal distribution with mean μ and precision τ. You want to infer the posterior distribution of μ. Your prior for μ also follows a normal distribution $\mu \sim N(\mu_0, \tau_0)$ with hyperparameters $\mu_0 = 10$ and $\tau_0 = 2$. Now you observe a sample of size $n = 10$, mean $\mu = 20$ and precision $\tau = 1/5$. What is your posterior μ_{post}?

(2 points)

	a	b	c	d	e
6:	other value	15	20	11	12

Question 7: What is your posterior τ_{post}

(2 points)

	a	b	c	d	e
7:	other value	6	1/2	2	4

Question 8: In a linear regression model you explain growth as a function of public expenditure and a country dummy. You have n country dummies in your model. For each dummy you test separately the Null hypothesis that the coefficient for this dummy is zero. Your level of significance is α. If the coefficients in the population are, indeed, all zero, how probable is it to find one or more statistically significant dummies in your sample?

(3 points)

	a	b	c	d	e
8:	other value	$n \cdot \alpha$	α^n	$1-(1-\alpha)^n$	$1-(1-\alpha^n)$

Question 9: Y is a binary variable which can be either 1 or 0. $\mathcal{L}$ is the logistic distribution function. You estimate the following model:

$$\Pr(Y = 1|X) = \mathcal{L}(\beta_0 + \beta_1 X)$$

Your estimate is $\beta_0 = 1$ and $\beta_1 = 0$. What are the estimated odds for $\Pr(Y = 1)/\Pr(Y = 0)$?

(2 points)

	a	b	c	d	e
9:	other value	$\frac{e}{1+e}$	$\log(X)$	e^{-1}	e

Question 10: Y is a binary variable which can be either 1 or 0. Φ is the cumulative distribution function of the normal distribution. ϕ is the density function of the normal distribution with $\phi(x) = d\Phi(x)/dx$. You estimate the following model:

$$\Pr(Y = 1|X) = \Phi(\beta_0 + \beta_1 X)$$

Your estimate is $\beta_0 = -1$ and $\beta_1 = 2$. What is the marginal effect of X on $P(Y = 1)$?

(2 points)

	a	b	c	d	e
10:	other value	2	$2X$	$2\phi(-1+2X)$	$\Phi(-1+2X)$

Question 11: Estimating the effect of x on y with different data but with the same model you obtain the following output from `glm`:

```
                 Estimate Std. Error z value Pr(>|z|)
     (Intercept)       -1          1      -1   0.3174
               x        4          2       2   0.0455
```

Which statements are correct?

(more than one answer possible, 10 points)

11a: The effect of x on y is significant.

11b: The marginal effect of x on $P(y=1)/P(y=0)$ is e^4.

11c: The marginal effect of x on y is 4.

11d: The marginal effect of x on $P(y=1)$ is $4\phi(4x-1)$.

11e: The marginal effect of x on $P(y=1)/P(y=0)$ is e^{4x-1}.

Question 12: You use JAGS to estimate the following model:

```
for (i in 1:length(y)) { y[i] ~ dpois(p[i]) ;
    p[i] <- exp(b[1]+b[2]*x[i]) }
for(k in 1:2) { b[k]~dnorm (0,.0001) }
```

Let $\Pr(\sigma)$ denote the probability that σ is true. Which statement is correct?

(3 points)

12a: None of the following answers are correct.

12b: $\Pr(\texttt{y[i]}=k)=\lambda^k e^{-\lambda}/k!$ with $\lambda=\texttt{b[1]+x[i]*b[2]}$.

12c: $\Pr(\texttt{y[i]}=k)=\lambda^k e^{-\lambda}/k!$ with $k=\texttt{b[1]+x[i]*b[2]}$.

12d: $\Pr(\texttt{y[i]}=k)=\lambda^k e^{-\lambda}/k!$ with $\lambda=e^{\texttt{b[1]+x[i]*b[2]}}$.

12e: $\Pr(\texttt{y[i]}\geq k)=\lambda^k e^{-\lambda}/k!$ with $k=e^{\texttt{b[1]+x[i]*b[2]}}$.

Question 13: You use JAGS to estimate the following model:

```
for (i in 1:length(y)) {
  y[i] ~ dnegbin(p[i],theta)
  p[i] <- theta/(theta+mu[i])
  theta ~ dexp(0.01)
  mu[i] <- exp(b[0]+b[1]*x[i])
}
for(k in 1:2) { b[k] ~ dnorm(0,.0001) }
```

What type of dependent variable does one typically estimate with this kind of model?

(2 points)

13a: None of the following answers are correct.

13b: A variable which can take negative as well as positive values.

13c: A variable which can take only two possible values, e.g. employed and unemployed.

13d: A variable which can take any value larger than one.

13e: A count variable.

Question 14: You estimate the following multinomial logit model:

$$\eta_a = x'\beta_a + \xi_a$$
$$\eta_b = x'\beta_b + \xi_b$$
$$\eta_c = x'\beta_c + \xi_c$$

The decision maker chooses alternative $k \in \{a,b,c\}$ if $\eta_k \geq \eta_j$ for all $j \in \{a,b,c\}$.
You obtain the following output in **R**:

```
multinom(formula = y~x1+x2+x3-1)
Coefficients:
  x1 x2 x3
b  2  4  1
c  3  2  3
```

With the same data you repeat the estimation, now using a different outcome as a reference. You obtain the following output (α_{ij} are placeholders):

```
multinom(formula = y~x1+x2-1)
Coefficients:
   x1  x2  x3
a α11 α12 α13
c α21 α22 α23
```

What do you expect for α_{12}?

(3 points)

	a	b	c	d	e
14:	other value	1	4	-4	-2

Question 15: What do you expect for α_{21}?

(3 points)

	a	b	c	d	e
15:	other value	3	-2	0	1

Question 16:
Consider the following output from **R**:

```
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ x + (1 + x | g)
Random effects:
 Groups   Name        Std.Dev.
 g        (Intercept)  1.0544
          x            0.9773
 Residual              1.1532
Number of obs: 15, groups: g, 4
Fixed effects:
            Estimate Std. Error t value
 (Intercept)  -2.3270     0.6155  -3.781
           x   0.9563     0.6586   1.452
```

(more than one answer possible, 5 points)

16a: None of the following answers are correct.

16b: The model includes a random effect only for the intercept.

16c: The model includes a random effect both for slope and intercept.

16d: The model can be written as $y_{it}=\beta_0+\nu_i+\beta_1 x_{it}+\epsilon_{it}$

16e: The model can be written as
$y_{it}=\beta_0+\nu_i+(\beta_1+\nu_i')x_{it}+\epsilon_{it}$

Question 17: You use the following command in **R** to estimate a model with mixed effects:

```
lmer(y~x+(1+x|g))
```

To represent this model in JAGS you write the following program (α is a placeholder):

```
for(k in 1:length(y)){
    y[k]~dnorm(beta[1]+beta[2]*x[k]+α,tau[1])
}
for(k in 1:3){
beta[k]~dnorm(0,.0001)
tau[k]~dgamma(.01,.01)
for(j in 1:max(g)) {nu[k,j]~dnorm(0,tau[j]) }
}
```

y and x are two $G \times T$ long vectors denoting the dependent variable y and the independent x. g is a $G \times T$ long vector which denotes the individual group.
What should be the value of α ?

(3 points)

17a: None of the following answers are correct.

17b: `x[k]*nu[g[k]]`

17c: `nu[g[k]]`

17d: `nu[1,k[g]]+x[k]*nu[2,k[g]]`

17e: `nu[1,g[k]]+x[k]*nu[2,g[k]]`

Question 18: You want to estimate the following relationship:

$$\mathtt{Y} = \beta_0 + \beta_1 \mathtt{X1} + \epsilon$$

You fear that `X1` and ϵ are correlated. You decide to use `Z1` as an instrument for `X1`. How can you estimate this problem with `ivreg`?

(3 points)

18a: None of the following answers are correct.

18b: `ivreg(Y~X1+Z1 | X1)`

18c: `ivreg(Y~Z1 | X1+Z1)`

18d: `ivreg(Y~Z1 | X1)`

18e: `ivreg(Y~X1 | Z1)`

Question 19: You still want to estimate the model

$$\mathtt{Y} = \beta_0 + \beta_1 \mathtt{X1} + \epsilon$$

still using `Z1` as an instrument. Consider the following model for JAGS (α and β are placeholders):

```
for(i in 1:length(y)) {
  α
  xH[i] <- g[1]+g[2]*z1[i];
  y[i] ~ dnorm(b[1]+β,tau[1])
}
for(k in 1:2) {
  b[k]~dnorm(0,.0001); g[k]~dnorm(0,.0001); tau[k] ~dexp(0.01)
}
```

What is the value of α if you want to estimate the model from the previous question? (3 points)

19a: other value

19b: `xH[i]~pnorm( x1[i], tau[2]);`

19c: `x1[i]~dnorm( xH[i], tau[2]);`

19d: `xH[i]~dnorm( x1[i], tau[2]);`

19e: `x1[i]~pnorm( xH[i], tau[2]);`

Question 20: What is the value of β for the previous model?

(3 points)

	a	b	c	d	e
20:	other value	`b[2]*xH[i]`	`b[2]*x1[i]`	`b[2]*z1[i]`	`b[2]*zH[i]`

Question 21: You want to explain Y as a linear function of either X_1 or X_2. You are unsure which model to use, either $Y = \beta_0 + \beta_1 X_1$ or $Y = \beta_0 + \beta_1 X_2$ You use the following model for model selection in JAGS (α is a placeholder):

```
for (i in 1:length(y)) {
  y[i]~dnorm(ifelse(equals(mI,1),α),tau[mI+1])
}
for (j in 1:4) { b[j]~dnorm(0,.001) }
for (j in 1:2) { tau[j]~dexp(0.01)}
mI~dbern(p)
p ~ dunif(0,1)
```

`y` contains your Y, `x1` and `x2` contain your X_1 and X_2, respectively. What should you fill in for α?

(3 points)

21a: None of the following answers are correct.

21b: `b[1]+b[2]*x1[i], b[1]+b[2]*x2[i]`

21c: `b[1]+b[2]*x1[i]+b[3]*x[2]`

21d: `y <- b[1]+b[2]*x1[i]`

21e: `b[1]+b[2]*x1[i], b[3]+b[4]*x2[i]`

Question 22:
In your output from the previous model you obtain a mean for `mI` of 0. What are the odds of the model $Y = \beta_0 + \beta_1 X_1$?

(2 points)

	a	b	c	d	e
22:	other value	0	$e/(1+e)$	1	1/2

7 MW24.1, February 2017 Final

Unless stated otherwise always use a significance level of 5%.

Question 1: You have estimated the following model in **R**:
`est<-glm(y~x,family=binomial(link=logit))`
Which type of model did you estimate?

(2 points)

1a: None of the following answers are correct.

1b: A negative binomial model

1c: A probit model

1d: A logistic model

1e: A linear model

Question 2: With the previous command you estimated two coefficients. Let us call them β_0 and β_x. Furthermore, `N` is a very large number. You issue the following commands in **R**:

```
B<-replicate(N,coef(glm(y~sample(x),
                 family=binomial))[2])
mean(B<coef(est)[2])
```

For which alternative hypothesis does the previous command provide a p-value?

(3 points)

	a	b	c	d	e
2:	other value	$\beta_x < 0$	$\beta_x > 0$	$\beta_x \neq 0$	$\beta_x = 0$

Question 3: How do you estimate the standard error of the estimated variance of `x` based on 1000 bootstrap replications?

(2 points)

3a: None of the following answers are correct.

3b: `sd(replicate(1000,var(sample(x,replace=FALSE))))`

3c: `var(replicate(1000,sd(sample(x,replace=TRUE))))`

3d: `var(replicate(1000,sd(sample(x,replace=FALSE))))`

3e: `sd(replicate(1000,var(sample(x,replace=TRUE))))`

Question 4: An event can have two possible outcomes, 0 or 1. You are interested in the probability p of obtaining a 1. You assume that p follows a Beta distribution. Your prior is that the parameters of the Beta distribution are $\alpha = \beta = 0$. You observe three times a 1 and no 0. What is your posterior for α and β?

(2 points)

	a	b	c	d	e
4:	other value	$\alpha = 3/2, \beta = 1$	$\alpha = 1, \beta = 1/3$	$\alpha = 3, \beta = 0$	$\alpha = \beta = 3$

Question 5: Later you observe three more times a 1 and four times 0. Given all your observations, what is now your posterior for α and β?

(2 points)

	a	b	c	d	e
5:	other value	$\alpha = \beta = 4/6$	$\alpha = 1, \beta = 0$	$\alpha = 6, \beta = 4$	$\alpha = \beta = 6/4$

Question 6: You use JAGS to estimate the following model:

```
for (i in 1:length(x)) { x[i]~dnorm(mu,tau) }
mu~dnorm (0,.0001);
tau~dexp(.01)
```

You obtain the following output:

```
     Lower95 Median Upper95  Mean   SD   psrf
mu    44.86  50.13   55.15 50.12 2.63 1.0001
```

Which statements are correct?

(more than one answer possible, 10 points)

6a: The p-value of this test is 50.12%

6b: `psrf` is very close to 1 which suggests that the samplers did converge.

6c: With probability 95% the posterior `x` is between 44.86 and 55.15.

6d: With probability 95% the posterior population mean of `x` is between 44.86 and 55.15.

6e: The posterior population mean of `x` has a standard deviation of 2.63.

Question 7: A researcher wants to find out whether a widget is good or bad. The Null hypothesis is that it is good. Good widgets always pass the researcher's test. Bad widgets pass the test with probability 10%. The researcher applies the following procedure: If the widget fails a first test, it is deemed bad. Otherwise the researcher carries out a second test. If the widget fails the second test, it is deemed bad. Otherwise it is deemed good. How probable is it that a good widget is deemed good?

(3 points)

	a	b	c	d	e
7:	other value	90%	99%	100%	81%

Question 8: Now the researcher assumes a prior probability of $1/2$ that the widget is good. The widget does pass a first test. What is the posterior probability that the widget is good?

(2 points)

	a	b	c	d	e
8:	other value	10/11	1	1/11	1/2

Question 9: With the same prior probabilities another widget fails the first test. What is now the posterior probability that the widget is good?

(2 points)

	a	b	c	d	e
9:	other value	1	0	1/11	9/19

Question 10: Y is a binary variable which can be either 1 or 0. $\mathcal{L}$ is the logistic distribution function. You estimate the following model:

$$\Pr(Y = 1|X) = \mathcal{L}(\beta_0 + \beta_1 X)$$

Your estimate is $\beta_0 = 1$ and $\beta_1 = -1$. What are the estimated odds for $\Pr(Y = 1)/\Pr(Y = 0)$?

(2 points)

	a	b	c	d	e
10:	other value	e^{1-X}	$\frac{e^{X-1}}{1+e^{X-1}}$	$\log(1 - X)$	$\mathcal{L}(X - 1)$

Question 11: Estimating the effect of x on y with different data but with the same model as in the previous question you obtain the following output from `glm`:

```
            Estimate Std. Error z value Pr(>|z|)
(Intercept)       -1          1      -1   0.3174
x                  4          2       2   0.0455
```

Which statements are correct?

(more than one answer possible, 10 points)

11a: The effect of x on y is significant.

11b: The marginal effect of x on $P(y = 1)$ is e^{4x-1}.

11c: The marginal effect of x on $P(y = 1)/P(y = 0)$ is e^{4x-1}.

11d: The marginal effect of x on $P(y = 1)/P(y = 0)$ is 4.

11e: The marginal effect of x on y is 4.

Question 12: You use JAGS to estimate the following model (α and β are placeholders):

```
for (i in 1:length(y)) { α ; β }
for(k in 1:2) {b[k]~dnorm (0,.0001) }
```

Assume that α has the value `y[i]~dpois(p[i])`
What type of dependent variable does one typically estimate with this kind of model?

(3 points)

12a: None of the following answers are correct.

12b: A variable which can take only two possible values, e.g. employed and unemployed.

12c: A variable which can take only values larger than one.

12d: A count variable.

12e: A variable which can take negative as well as positive values.

Question 13: Now you want to estimate a negative binomial model where y is a function of an independent variable x. Assume that in the previous specification β has the following value:

```
p[i] <- theta/(theta+mu[i])
theta ~ dexp(0.01)
mu[i] <- exp(b[0]+b[1]*x[i])
```

What should now be the value of α?

(2 points)

13a: None of the following answers are correct.

13b: `y<-dnegbin(p[i],theta)`

13c: `y[i]~dnegbin(p[i],theta)`

13d: `y[i]~dnegbin(p,theta)`

13e: `y~dnegbin(p[i],theta)`

Question 14: You estimate the following multinomial logit model:

$$\eta_a = x'\beta_a + \xi_a$$
$$\eta_b = x'\beta_b + \xi_b$$
$$\eta_c = x'\beta_c + \xi_c$$

The decision maker chooses alternative $k \in \{a,b,c\}$ if $\eta_k \geq \eta_j$ for all $j \in \{a,b,c\}$.
You obtain the following output in **R**:

```
multinom(formula = y~x1+x2-1)
Coefficients:
  x1 x2
b  1  4
c  0  5
```

With the same data you repeat the estimation, now using a different outcome as a reference. You obtain the following output (α_{ij} are placeholders):

```
multinom(formula = y~x1+x2-1)
Coefficients:
   x1   x2
a α11  α12
b α21  α22
```

What do you expect for α_{12}?

(3 points)

14:	a	b	c	d	e
	other value	-5	-1	0	1

What do you expect for α_{21}?

(3 points)

15:	a	b	c	d	e
	other value	-1	0	1	-4

Question 16:
Consider the following output from **R**:

```
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ x + (1 | g)
REML criterion at convergence: 79.4435
Random effects:
 Groups   Name        Std.Dev.
 g        (Intercept)  0.6237
 Residual              1.3024
Number of obs: 23, groups: g, 4
Fixed Effects:
 (Intercept)      x
      5.358 2.772
```

(more than one answer possible, 5 points)

16a: None of the following answers are correct.

16b: The model can be written as $y_{it} = \beta_0 + \beta_1 x_{it} + \nu_i + \epsilon_{it}$

16c: The model can be written as $y_{it} = \beta_0 + \beta_1 x_{it} + \nu_{it} + \epsilon_{it}$

16d: The model includes a random effect for the intercept.

16e: The model includes a random effect for x.

Question 17: Consider the following model in JAGS (α is a placeholder):

```
for(k in 1:length(y)){
    y[k]~dnorm(beta[1]+beta[2]*x[k]+nu[g[k]],tau[1])
}
α
for(k in 1:2){beta[k]~dnorm(0,.0001);tau[k]~dgamma(.01,.01)}
```

y and x are two $G \times T$ long vectors denoting the dependent variable y and the independent x. g is a $G \times T$ long vector which denotes the individual group.
What should be the value of α ?

(3 points)

17a: None of the following answers are correct.

17b: `for(j in 1:length(g)){x[i,j]~dnorm(0,tau[2]) }`

17c: `for(j in 1:max(g)){nu[j]~dnorm(0,tau[2]) }`

17d: `for(j in 1:length(g)){nu[i,j]~dnorm(0,tau[2]) }`

17e: `for(j in 1:max(g)){x[j]~dnorm(0,tau[2]) }`

Question 18: You want to estimate the following relationship:

$$\texttt{Y} = \beta_0 + \beta_1\texttt{X1} + \beta_2\texttt{X2} + \epsilon$$

You fear that X2 and ϵ might be correlated. You decide to use Z1 and Z2 as instruments for X2. How can you estimate this problem with `ivreg`?

(3 points)

18a: None of the following answers are correct.

18b: `ivreg(Y~X1 + X2 | X1 + Z1 + Z2)`

18c: `ivreg(Y~X1 + X2 | Z1 + Z2)`

18d: `ivreg(Y~X1 + X2 + Z1 + Z2 | X1 + Z1 + Z2)`

18e: `ivreg(Y~X2 | X1 + Z1 + Z2)`

Question 19: Consider the following model for JAGS (α and β are placeholders):

```
for(i in 1:length(y)) {
  x2[i]~dnorm( xH[i], tau[2]);
  xH[i] α g[1]+g[2]*z1[i]+g[2]*z2[i];
  y[i] β dnorm(b[1]+b[2]*x1[i]+b[3]*xH[i],tau[1])
}
for(k in 1:3) {
  b[k]~dnorm(0,.0001); g[k]~dnorm(0,.0001); tau[k] ~dexp(0.01)
}
```

What are the values of α and β if you want to estimate the model from the previous question? (3 points)

19:	a	b	c	d	e
	other value	α: ~ β: ~	α: <- β: <-	α: ~ β: <-	α: <- β: ~

Question 20: You use the following model for model selection in JAGS (α is a placeholder):

```
for (i in 1:length(y)) {
  y[i]~dnorm(ifelse(equals(mI,0),
      b[1]+b[2]*x1[i]+b[3]*x2[i]+b[4]*x1[i]*x2[i],
      b[5]+b[6]*x1[i]+b[7]*x2[i]),
    tau[mI+1])
}
for (j in 1:7) { b[j]~dnorm(0,.001) }
for (j in 1:2) { tau[j]~dexp(0.01)}
mI~dbern(p)
α
```

What should you fill in for α?

(2 points)

20:	a	b	c	d	e
	other value	p <- dnorm(0,1)	p <- dunif(0,1)	p ~ dunif(0,1)	p ~ dnorm(0,1)

Question 21: Which econometric models does this JAGS model compare?

(2 points)

21a: None of the following answers are correct.

21b: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$ against
$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon$

21c: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$ against
$Y = \beta_0 + \beta_1 X_1 * \beta_2 X_2 + \epsilon$

21d: $Y = \beta_0 + \beta_1 X_1 + \epsilon$ against
$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon$

21e: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$ against
$Y = \beta_0 + \beta_3 X_1 X_2 + \epsilon$

Question 22:
In your output from JAGS you obtain a 95%-credible interval for `mI` of $CI_{95\%} = [0, 1]$ and a mean for `mI` of 0.2. What are the odds of the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$? (3 points)

22:	a other value	b 5	c 4	d 1/4	e 1/5

---

8 MW24.1, April 2016 Final Resit

Question 1: You have estimated the model $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ and found parameter estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ as well as estimated residuals $\hat{\epsilon}_i$.
You want to use a bootstrap to estimate a standard deviation of $\hat{\beta}_1$. To do this, you draw a large number B of bootstrap replications $\hat{\beta}_1^{*k}$ and use their standard deviation $\text{se}_{k=1}^{B}(\hat{\beta}_1^{*k})$ as your estimator. You obtain these $\hat{\beta}_1^{*k}$ from estimating the following equation:

$$Y_i^{*k} = \beta_0^{*k} + \beta_1^{*k} X_i + \epsilon_i^k$$

What should you use for Y_i^{*k}? (in the following the function "`sample`(x)" is a function that returns a sample of x in a random order but with the same size as x.) (3 points)

1a: None of the following answers are correct.

1b: $Y_i^{*k} = \hat{\beta}_0 + \hat{\beta}_1 \texttt{sample}(X_i) + \hat{\epsilon}_i$

1c: $Y_i^{*k} = \texttt{sample}(\hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\epsilon}_i)$

1d: $Y_i^{*k} = \hat{\beta}_0 + \hat{\beta}_1 X_i + \texttt{sample}(\hat{\epsilon}_i)$

1e: $Y_i^{*k} = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\epsilon}_i$

Question 2: Assume that in the previous question you have found that never $\hat{\beta}_1^{*k} = 0$ and in 98% of your bootstraps $\hat{\beta}_1^{*k} > 0$. What is the p-value for the two-sided test of H_0 that $\beta_1 = 0$? (2 points)

2:	a other value	b .96	c .98	d .02	e .04

Question 3: There are two possible states of the world: A and B. Your Null hypothesis is that you are in state A. Depending on the true state your test generates outcomes a and b with the following probabilities:

		outcome	
		a	b
state	A	.9	.1
	B	.5	.5

Successive test outcomes are independent of each other. You use the following protocol: You apply your test once. If the outcome is b, you stop and reject your Null hypothesis. If the outcome is a, you apply the test again. If this time the outcome is b, you stop and reject your Null hypothesis. Otherwise you stop and do not reject. How probable is it that you reject if you are in state A? (3 points)

3:	a other value	b .45	c .01	d .1	e .19

Question 4: Consider the above situation again. Now your prior is that with equal probability you are in state A or B. You apply your test once and you observe the outcome b. What is your posterior probability to be in state B? (2 points)

4:	a other value	b 1/10	c 1/6	d 9/14	e 5/6

Question 5: What must (in the previous question) your prior $\Pr(A)$ be, such that after the test (where you observe b) you find A and B equally probable? (3 points)

5:	a other value	b 1/6	c 1/3	d 5/6	e 9/10

Question 6: An event can have two possible outcomes, success or failure. You are interested in the probability p of a success. You assume that p follows a Beta distribution. Your prior is that the parameters of the Beta distribution $\alpha = \beta = 1$. Now you observe three failures and no success. What is your posterior for α and β? (2 points)

6:	a other value	b $\alpha = 4, \beta = 1$	c $\alpha = 0, \beta = 3$	d $\alpha = \beta = 1$	e $\alpha = 1, \beta = 4$

Question 7: You use JAGS to estimate the following model:

```
for (k in 1:length(x)) { x[k] ~ dnorm(beta,tau) }
beta ~ dnorm (0,.0001);
tau ~ dexp(.01)
```

You obtain the following output:

```
       Lower95 Median Upper95 Mean  SD   psrf
beta   12.3     14      16     14  1.2 1.0002
```

Which statements are correct? (more than one answer possible, 10 points)

7a: None of the following answers are correct.

7b: The posterior distribution of the true mean of `x` has a standard deviation of 1.2

7c: The p-value of this test is 1.0002.

7d: With probability 95% `x` is between 12.3 and 16.

7e: With probability 95% the true mean of `x` is between 12.3 and 16.

Question 8: You want to measure the effect of a continuous effort `x` on success `y` where `y=1` in case of a success and `y=0` in case of a non-success. With R which command yields the desired result? (2 points)

8a: None of the following answers are correct.

8b: `glm(y ~ x)`

8c: `glm(y ~ x,family=binomial(link=probit))`

8d: `glm(y ~ x,family=poisson(link=logit))`

8e: `glm(y ~ logit(x)`

Question 9: Y is a binary variable which can be either 1 or 0. $\mathcal{L}$ is the logistic function. You estimate the following model:

$$\Pr(Y = 1|X) = \mathcal{L}(\beta_0 + \beta_1 X)$$

Your estimate is $\hat{\beta}_0 = 0$ and $\hat{\beta}_1 = 1$. What are the estimated odds for $\Pr(Y = 1)/\Pr(Y = 0)$? (3 points)

9:	a other value	b e^X	c $\mathcal{L}(X)$	d $\frac{X}{1-X}$	e X

Question 10: You use JAGS to estimate a probit model. Consider the following model (α and β are placeholders):

```
for (i in 1:length(y)) { α ; β }
for(k in 1:2) {b[k] ~ dnorm (0,.0001) }
```

Assume that α has the value `y[i] ~ dbern(p[i])`
What should be the value of β: (3 points)

10a: None of the following answers are correct.

10b: `p[i] <- pnorm(b[1]+b[2]*x[i],0,1)`

10c: `p[i] <- dnorm(b[1]+b[2]*x[i],0,1)`

10d: `p[i] <- plogis(b[1]+b[2]*x[i],0,1)`

10e: `p[i] <- dlogis(b[1]+b[2]*x[i],0,1)`

Question 11: Now `y` is a count variable. You want to use a negative binomial model. What should be the value of α? (2 points)

11a: None of the following answers are correct.

11b: `y[i] <- dpois(p[i],r)`

11c: `y[i] <- dnegbin(p[i],r)`

11d: `y[i] ~ dnegbin(p[i],r)`

11e: `y[i] ~ dpois(p[i],r)`

Question 12: Which of the following statements are correct? (more than one answer possible, 5 points)

12a: The negative binomial model is a special case of the Poisson model

12b: The Poisson model is a special case of the negative binomial model

12c: The negative binomial model allows us to model overdispersion

12d: The Poisson model allows us to model underdispersion

12e: The Poisson model is more appropriate than the negative binomial if we have too many zeroes in our data

Question 13: You estimate the following multinomial logit model:

$$\eta_1 = x'\beta_1 + \xi_1$$
$$\eta_2 = x'\beta_2 + \xi_2$$
$$\eta_3 = x'\beta_3 + \xi_3$$

The decision maker chooses alternative k if $\eta_k \geq \eta_j$ for all j. You do not know the variance of ξ_1, ξ_2, ξ_3. You expect the following:

$$(\beta_1, \beta_2, \beta_3) = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 2 & 1 \end{pmatrix}$$

You obtain the following output in R (α_1 and α_2 are placeholders):

```
multinom(formula = y ~ x1+x2-1)
Coefficients:
   x1   x2
2   2    0
3  α1   α2
```

What do you expect for α_1? (3 points)

13:	a other value	b 0	c 1	d 2	e −2

Question 14: What do you expect for α_2? (3 points)

14:	a other value	b 0	c 1	d 2	e −2

Question 15: You estimate the following mixed effects model

$$y_{it} = \beta_1 + \beta_2 x_{it} + \nu_i + \nu'_j + \epsilon_{it}$$

where $i \in \{1, \ldots, I\}$ denotes the individual, $j \in \{1, \ldots, J\}$ denotes a group, and $t \in \{1, \ldots, T\}$ denotes time. ϵ_{it} is the residual, ν_i is the individual specific random effect, ν'_j is the group specific random effect. The vector `j[i]` tells you to which group individual `i` belongs. You use the following model in JAGS (α is a placeholder):

```
for(k in 1:length(y)){
   y[k] ~ dnorm( beta[1] + beta[2]*x[k] + α, tau[1] )
}
for(k in 1:max(i)) { nu[k] ~ dnorm(0,tau[2]) }
for(k in 1:max(j)) { nu2[k] ~ dnorm(0,tau[3]) }
for(k in 1:2) { beta[k] ~ dnorm(0,.0001) }
for(k in 1:3) { tau[k] ~ dgamma(.01,.01) }
```

Here `y` and `x` are two $I \times T$ long vectors denoting the dependent variable y and the independent x. `i` is a vector with length $I \times T$ denoting to which individual i each observation k belongs. `j` is a vector of length I denoting to which group j each individual i belongs.
What should be the value of α ?

(4 points)

15a: None of the following answers are correct.

15b: `nu[i[k]]] + nu2[i[j[k]]]`

15c: `nu[k] + nu2[k]`

15d: `nu[i[k]] + nu2[j[i[k]]]`

15e: `nu[j[i[k]]] + nu2[j[i[k]]]`

Question 16: You estimated a mixed effects model with `lmer` and you have stored the result in the variable `mer`. You have also estimated the same model, though without the random effects, with `lm` and stored the result in the variable `ols`. To find out whether the mixed effects model is justified, you run the following approximate permuatation test:

```
lls <- replicate(500,{
   y1 <- simulate(ols)[[1]]
   l.ols <- logLik(lm (y1 ~ x))[1]
   l.mer <- logLik(refit(mer,y1))[1]
   2*(l.mer-l.ols)
})
mean(lls < 2*(logLik(mer)-logLik(ols))[1])
```

The last statement returns the value 0.98. Your Null hypothesis is that there is no random effect. What is the p-value? (3 points)

16:	a other value	b .02	c .04	d .96	e .98

Question 17: The variables x1, x2, and x3 are independent and from a continuous distribution. Which of the following instrumental variable models are exactly identified?
(more than one answer possible, 5 points)

17a: `ivreg(y~x1|x1)`

17b: `ivreg(y~x1+x3|x2)`

17c: `ivreg(y~x1+x3|x2+I(x2^2))`

17d: `ivreg(y~x1|x2+x3)`

17e: `ivreg(y~x1|x2)`

Question 18: You want to estimate an equivalent of the following instrumental variables model

`ivreg(y~x1|x2+x3)`

Consider the following model for JAGS (α and β are a placeholders):

```
for(i in 1:length(y)) {
  x1[i] ~ dnorm( x1H[i], tau[2])
  x1H[i] <- α
  y[i] ~ dnorm(β, tau[1])
}
for(k in 1:3) {
  b[k] ~ dnorm(0,.0001)
  c[k] ~ dnorm(0,.0001)
  tau[k] ~ dgamma(0.01,.01)
}
```

What is the value of β? (3 points)

18a: None of the following answers are correct.

18b: `x1[i]`

18c: `x1H[i]`

18d: `b[1]+b[2]*x1H[i]`

18e: `b[1]+b[2]*x1[i]`

Question 19: What is the value of α? (3 points)

19a: None of the following answers are correct.

19b: `c[1]*x2+c[2]*x3`

19c: `c[1]+c[2]*x2[i]+c[3]*x3[i]`

19d: `c[1]+c[2]*x[2]+c[3]*x[3]`

19e: `x2+x3`

Question 20: Your sample contains a random variable y which is drawn from one of two possible distributions. Model A assumes a normal distribution with mean 0 and precision 1. Model B assumes a normal distribution with mean 3 and precision 1. You use JAGS to determine the posterior probabilities. Consider the following model:

```
for (i in 1:length(y)) {
  y[i] ~ α
}
j ~ dbern(1/3)
```

What is the value of α? (3 points)

20a: None of the following answers are correct.

20b: `dnorm(ifelse(equals(j,1),0,3),1)`

20c: `ifelse(dnorm(equals(j,1),0,3),1)`

20d: `dnorm(equals(ifelse(j,1),0,3),1)`

20e: `equals(dnorm(ifelse(j,1),0,3),1)`

Question 21: What was your prior probability of Model A in the above model? (2 points)

21:	a other value	b 1/2	c 2/3	d 1	e 1/3

---

9 MW24.1, February 2016 Final

Unless stated otherwise always use a significance level of 5%.

Question 1: How do you estimate the standard error of the estimated median of x based on 100 bootstrap replications? (3 points)

1a: None of the following answers are correct.

1b: `sd(replicate(100,median(sample(x,replace=FALSE))))`

1c: `sd(sample(100,median(replicate(x,replace=FALSE))))`

1d: `sd(replicate(100,median(sample(x,replace=TRUE))))`

1e: `sd(sample(100,median(replicate(x,replace=TRUE))))`

Question 2: The variable x contains income of a worker, the variable t describes whether the worker did (t=1) or did not (t=0) participate in a qualification program. The function `t.test(x~t)$statistic` yields a t-statistic of 1.4 for a comparison of x for the two groups which are described by t. Since your sample is small you are not sure that the t-statistic really follows a t-distribution. Instead you use an approximate permutation test. How do you get the p-value for a two-sided test of the Null hypothesis that both groups have on average the same income? (3 points)

2a: None of the following answers are correct.

2b: `max(1.4>replicate(1000,t.test(x~sample(t))$statistic))`

2c: `mean(1.4<replicate(1000,abs(t.test(x~sample(t))$statistic)))`

2d: `mean(1.4<replicate(1000,t.test(x~sample(t))$statistic))`

2e: `max(1.4>replicate(1000,abs(t.test(x~sample(t))$statistic)))`

Question 3: A researcher wants to find out whether a widget is good or bad. The Null hypothesis is that it is good. Good widgets fail a given test with probability $1/10$. Bad widgets always fail the test. The researcher applies the following procedure: If the widget fails a first test, it is deemed bad. Otherwise the researcher carries out a second test. If the widget fails the second test, it is deemed bad. Otherwise it is deemed good. How probable is it that a good widget is deemed good? (3 points)

3:	a other value	b $8/10$	c $(9/10)^2$	d $(9/10)$	e $19/20$

Question 4: Now the researcher assumes a prior probability of 1/2 that the widget is good. The widget does pass a first test. What is the posterior probability that the widget is good?

(2 points)

4:	a other value	b 1/2	c 0.9	d 1	e 0

Question 5: With the same prior probabilities another widget does not pass a first test. What is now the posterior probability that the widget is good?

(2 points)

5:	a other value	b 9/19	c 1/11	d 0	e 1/2

Question 6: You use JAGS to estimate the following model:

```
for (i in 1:length(x)) { x[i]~dnorm(mu,tau) }
mu ~ dnorm (0,.0001);
tau ~ dexp(.01)
```

You obtain the following output:

```
   Lower95 Median Upper95 Mean  SD  psrf
mu  33.7     35    36.3    35  0.7 1.0001
```

Which conclusions are correct?

(more than one answer possible, 10 points)

6a: The p-value of this test is 0.7

6b: `x` has a standard deviation of 0.7

6c: `psrf` is very close to 1 which suggests that the samplers did converge.

6d: With probability 95% `x` is between 33.7 and 36.3.

6e: With probability 95% the true mean of `x` is between 33.7 and 36.3.

Question 7: You want to measure the effect of a continuous effort `x` on success `y` where `y=1` in case of a success and `y=0` in case of a non-success. With R which command yields the desired result?

(2 points)

7a: None of the following answers are correct.

7b: `glm(y ~ x)`

7c: `glm(y ~ x,family=binomial(link=logit))`

7d: `glm(y ~ x,family=poisson(link=log))`

7e: `glm(y ~ logit(x))`

Question 8: You obtain the following output from a `glm` model:

```
Call:  glm(formula=y~x,family=binomial(link=logit))
               Estimate Std. Error z value Pr(>|z|)
(Intercept)       -10      40000   -0.00     1.00
          x       120     200000    0.00     1.00
```

Which statements are correct?

(more than one answer possible, 10 points)

8a: The effect of `x` on `y` is not significant.

8b: The marginal effect of `x` on $\Pr(y = 1)$ is $e^{120x-10}$.

8c: The marginal effect of `x` on $\Pr(y = 1)/\Pr(y = 0)$ is $e^{120x-10}$.

8d: The marginal effect of `x` on $\Pr(y = 1)/\Pr(y = 0)$ is 120.

8e: The marginal effect of `x` on `y` is 120.

Question 9: You use JAGS to estimate a logit model. Consider the following model (α and β are placeholders):

```
for (i in 1:length(y)) { α ; β }
for(k in 1:2) {b[k] ~ dnorm (0,.0001) }
```

Assume that α has the value `y[i] ~ dbern(p[i])`

What should be the value of β:

(3 points)

9a: None of the following answers are correct.

9b: `p[i] <- plogis(b[1]+b[2]*x[i],0,1)`

9c: `p[i] ~ plogis(b[1]+b[2]*x[i],0,1)`

9d: `y[i] ~ plogis(b[1]+b[2]*x[i],0,1)`

9e: `y[i] <- plogis(b[1]+b[2]*x[i],0,1)`

Question 10: Now `y` is a count variable. You want to use a Poisson model. What should be the value of α?

(2 points)

10a: None of the following answers are correct.

10b: `y[i] ~ dpois(p[i])`

10c: `y[i] <- dpois(p[i])`

10d: `y[i] ~ dlogis(p[i])`

10e: `y[i] ~ dprobit(p[i])`

Question 11: What should in this case be the value of β?

(2 points)

11a: None of the following answers are correct.

11b: `p[i] <- exp(b[1]+b[2]*x[i])`

11c: `p[i] ~ exp(b[1]+b[2]*x[i])`

11d: `p[i] <- log(b[1]+b[2]*x[i])`

11e: `p[i] ~ b[1]+b[2]*x[i]`

Question 12: You estimate the following multinomial logit model:

$$\eta_1 = x'\beta_1 + \xi_1$$
$$\eta_2 = x'\beta_2 + \xi_2$$
$$\eta_3 = x'\beta_3 + \xi_3$$

The decision maker chooses alternative k if $\eta_k \geq \eta_j$ for all j. You expect the following:

$$(\beta_1, \beta_2, \beta_3) = \begin{pmatrix} 3 & 4 & 1 \\ 2 & 0 & 6 \end{pmatrix}$$

You obtain the following output in R (α_1 and α_2 are placeholders):

```
multinom(formula = y ~ x1+x2-1)
Coefficients:
  x1 x2
2  1 -2
3 α1 α2
```

What do you expect for α_1?

(3 points)

12:	a other value	b 0	c 1	d 2	e -2

What do you expect for α_2?

(3 points)

13:	a other value	b 4	c 6	d 0	e 2

Question 14: You estimate the following mixed effects model

$$y_{i,t} = \beta_1 + \beta_2 x_{i,t} + \nu_i + \epsilon_{i,t}$$

where $i \in \{1, \ldots, I\}$ denotes the individual, and $t \in \{1, \ldots, T\}$ denotes time. $\epsilon_{i,t}$ is the residual, ν_i is the individual specific random effect. You use the following model in JAGS (α is a placeholder):

```
for(k in 1:length(y)){ α }
for(j in 1:max(i)){nu[j]~dnorm(0,tau[2])}
for(k in 1:2){beta[k]~dnorm(0,.0001);tau[k]~dgamma(.01,.01)}
```

Here y and x are two $I \times T$ long vectors denoting the dependent variable y and the independent x. i is a $I \times T$ long vector which denotes the individual i.
What should be the value of α ?

(3 points)

14a: None of the following answers are correct.

14b: `y[k] ~ dnorm(beta[1]+beta[2]*x[k]+i[k],tau[1])`

14c: `y[k] ~ dnorm(beta[1]+beta[2]*x[k]+nu[k],tau[1])`

14d: `y[k] ~ dnorm(beta[1]+beta[2]*x[k]+nu[i[k]],tau[1])`

14e: `y[k] ~ dnorm(beta[1]+beta[2]*x[k]+i[nu[k]],tau[1])`

Question 15:
You use `lmer` from the `lme4` library to estimate a mixed effects model `m.mer`. You also use `lm` to estimate a simple OLS model `m.ols`. A comparison of the two models with the help of `anova` yields the following result:

	Df	AIC	BIC	logLik	χ^2	Df	Pr(> χ^2)
m.ols	3	319	327	-156			
m.mer	4	317	327	-154	4.21	1	0.0401

Assume a level of significance of 5%. Which of the following statements is true?

(more than one answer possible, 10 points)

15a: The random effect is significant.

15b: The method used by `anova` here is often too conservative. The true p-value may well be smaller than 0.0401.

15c: The method used by `anova` here is often anti-conservative. The true p-value may well be larger than 0.0401.

15d: It is possible to obtain a better estimate for the p-value of this test with the help of a bootstrap.

15e: `anova` does a Likelihood ratio test here.

Question 16: You want to estimate the equation $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon$. You fear, however, that X and ϵ may be correlated. You decide to use Z1 as an instrument. You estimate the following regression:

```
ivreg(Y ~ X + I(X^2) | I(X^2) + Z1))
```

Which statements are true?

(more than one answer possible, 10 points)

16a: Z1 can only be good instrument for X if Z1 is not correlated with ϵ.

16b: The above approach is underidentified since only one instrument Z1 is used as an instrument for both X and X^2.

16c: The above approach is exactly identified since two instruments (Z1 and X^2) are used as instruments for two variables X and X^2.

16d: If Z1 is independent of ϵ, the above approach does provide a good instrument for X and for X^2.

16e: Z1 is only a good instrument if it is not correlated with X.

Question 17: Consider the following model for JAGS:

```
for(i in 1:length(y)) {
    x[i] ~ dnorm( xH[i], tau[2])
    xH[i] <- g[1] + g[2] * z[i]
    y[i] ~ dnorm(b[1] + b[2] * xH[i], tau[1])
}
for(k in 1:2) {
    b[k] ~ dnorm(0,.0001)
    g[k] ~ dnorm(0,.0001)
    tau[k] ~ dgamma(0.01,.01)
}
```

What is the equivalent of this model with `ivreg`?

(3 points)

17a: None of the following answers are correct.

17b: `ivreg(y ~ x + xH | z))`

17c: `ivreg(y ~ x | g1 + g2))`

17d: `ivreg(y ~ x + I(x^2) | I(x^2) + z))`

17e: `ivreg(y ~ x | z))`

Question 18: You estimate the following equation:

$$Y = \sum_{i=0}^{k} \beta_i X^i + \epsilon$$

The following table shows the AIC of the estimated model for different values of k:

k	1	2	3	4	5	6	7
AIC	192.78	165.21	147.83	90.07	78.05	5.93	7.76

Which value of k should you choose, based on the AIC?

(2 points)

	a	b	c	d	e
18:	other value	6	1	2	4

Question 19: You use the following model for model selection in JAGS (α is a placeholder):

```
for (i in 1:length(y)) {
   y[i] ~ dnorm(ifelse(equals(mI,0),
      b[1]+b[2]*x1[i],
      b[3]+b[4]*x2[i]),
     tau[mI+1])
}
for (j in 1:4) { b[j] ~ dnorm(0,.001) }
for (j in 1:2) { tau[j] ~ dexp(0.01)}
mI ~ α
p ~ dunif(0,1)
```

What should you fill in for α?

(3 points)

	a	b	c	d	e
19:	other value	`dpois(p)`	`dbern(p)`	`p`	`exp(p)`

Question 20: Which econometric models does this JAGS model compare?

(2 points)

20a: None of the following answers are correct.

20b: $Y = \beta_0 + \beta_1 X_1 + \epsilon$ against $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$

20c: $Y = \beta_0 + \beta_1 X_1 + \epsilon$ against $Y = \beta_0 + \beta_1 X_1^2 + \epsilon$

20d: $Y = \beta_0 + \beta_1 X_1 + \epsilon$ against $Y = \beta_0 + \epsilon$

20e: $Y = \beta_0 + \beta_1 X_1 + \epsilon$ against $Y = \beta_0 + \beta_1 X_2 + \epsilon$

Question 21: In your output from JAGS you obtain a 95%-credible interval for `mI` of $CI_{95\%} = [0, 1]$ and a mean for `mI` of 0.3. What can you conclude? (3 points)

21a: None of the following answers are correct.

21b: The probability of the model $Y = \beta_0 + \beta_1 X_1 + \epsilon$ is 0.5.

21c: Since the credible interval for `mI` includes the extremes 0 and 1 one can not make a statement how probable the models are.

21d: The probability of the model $Y = \beta_0 + \beta_1 X_1 + \epsilon$ is 0.3.

21e: The probability of the model $Y = \beta_0 + \beta_1 X_1 + \epsilon$ is 0.7.

---

10 MW24.1, December 2018 Midterm Resit

Question 1: According to your (two-sided) Null hypothesis the population mean of a normally distributed random variable is 2. In a sample with 7 observations you find a sample standard deviation of 3 and a sample mean of -2. Which of the following statements tells you whether you can reject this Null hypotheses at a level of significance of 10%? (3 points)

1a: None of the following answers are correct.

1b: `4 > qt(.05,df=6)*3/sqrt(7)`

1c: `-4 < qt(.05,df=6)*3/sqrt(7)`

1d: `-4 > qt(.05,df=6)*3/sqrt(7)`

1e: `4 < qt(.05,df=6)*3/sqrt(7)`

Question 2: How can one obtain a p-value for the above hypothesis test? (2 points)

2a: None of the following answers are correct.

2b: `1-2*pt(4/3*sqrt(7),df=6)`

2c: `2*pt(4/3*sqrt(7),df=6)`

2d: `pt(4/7*sqrt(7),df=6)`

2e: `2-pt(4/7*sqrt(7),df=6)`

Question 3: A variable `x` contains a sample of a random variable X. Consider the following output from R:

```
    One Sample t-test
data: x
t = 3.4244, df = 10, p-value = 0.0065
alternative hypothesis: true mean is not equal to 1
95 percent confidence interval:
 1.450165 3.127171
sample estimates:
mean of x
 2.288668
```

Which of the following statements is correct? (2 points)

3a: None of the following answers are correct.

3b: The probability that $E(X) = 1$ is equal to 0.0065.

3c: If $E(X) = 1$, then the probability to get a sample like `x` or one that is even more adverse to our Null Hypothesis is equal to 0.0065.

3d: If $X = 1$, then the probability to get a sample like `x` or one that is even more adverse to our Null Hypothesis is equal to 0.0065.

3e: The probability that $X = 1$ is equal to 0.0065.

Question 4: You use the `lm` function of R to estimate the relationship $Y_i = \beta_0 + \beta_1 X_i + u_i$. The result of your estimation is stored in `est`. Consider the following output from `confint(est)`:

```
                   2.5 %   97.5 %
(Intercept) -205.38234 518.4940
x            -97.75741 221.6415
```

Based on a level of significance of 5% and with a two sided test, which of the following Null hypotheses can be rejected? (more than one answer possible, 5 points)

4:	a	b	c	d	e
	$\beta_0 = 2$	$\beta_0 = 300$	$\beta_0 = -100$	$\beta_0 = 600$	$\beta_0 = 400$

Question 5: With a different sample, you estimate again $Y_i = \beta_0 + \beta_1 X_i + u_i$. As before, you store your estimation result in `est`. Consider the following output from a linear hypothesis test:

```
Hypothesis: x = 5
Model 1: restricted model
Model 2: y ~ x
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     14 153197
2     13 128977  1     24220 2.4412 0.1422
```

Your significance level is 1%. Which of the following statements are correct? (3 points)

5a: None of the following answers are correct.

5b: The probability that the alternative hypothesis is correct is 0.1422.

5c: The Null hypothesis of this test was that $\bar{X} = 5$.

5d: The Null hypothesis is rejected.

5e: The probability that the Null hypothesis is correct is 0.1422.

Question 6: In a sample with 12 observations of a normally distributed random variable one finds a sample standard deviation of $\sqrt{12}$ and a sample mean of -2. How does one determine the width of a 99%-confidence interval around the sample mean? (2 points)

6a: None of the following answers are correct.

6b: `qt(.005,df=11)`

6c: `-2*qt(.005,df=11)`

6d: `2*pt(.01,df=11)*sqrt12`

6e: `2*pt(.01,df=11)*12`

Question 7: A random variable X is distributed as follows: $\Pr(X = 1) = \theta$, $\Pr(X = 2) = 1 - \theta - \theta^2$, $\Pr(X = 3) = \theta^2$ with $0 \le \theta \le \frac{6}{10}$. Your sample is $\{1, 1, 2, 3\}$. What is the ML estimator for θ? (4 points)

	a	b	c	d	e
7:	other value	1/2	3/4	7/4	0

Question 8: A random variable follows a Normal distribution. The density function of the Normal distribution is $f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Your sample is $\{1, 2, 3\}$. What is the ML estimator for the parameter μ? (4 points)

	a	b	c	d	e
8:	other value	2	3	0	1

Question 9: Consider the relationship $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2$.
If $X_1 = 0$ you expect the marginal effect of X_2 on Y to be 2.
If $X_1 = 1$ you expect the marginal effect of X_2 on Y to be 4.
If $X_2 = 1$ you expect the marginal effect of X_1 on Y to be 2.
What do you expect for β_1? (1 point)

	a	b	c	d	e
9:	other value	4	0	1	2

Question 10: What do you expect for β_2? (1 point)

	a	b	c	d	e
10:	other value	0	1	2	4

Question 11: What do you expect for β_3? (1 point)

	a	b	c	d	e
11:	other value	1	2	4	0

Question 12: Consider the following regression model:

$$Y = \beta_0 + \beta_1 \log X_1 + \beta_2 \log X_2 + u$$

Your estimate for β_2 is 0.3. Which of the following statements is (approximately) true? (3 points)

12a: None of the following answers are correct.

12b: An increase of X_2 by 1% leads to an increase of Y by 0.3 units.

12c: An increase of X_2 by 0.3 units leads to a 1% increase of Y.

12d: An increase of X_2 by 30% leads to an increase of Y by one unit.

12e: An increase of X_2 of one unit leads to an increase of Y by 30%.

Question 13: You explain a dependent variable Y as a function of an independent X. You expect that, independent of the level of X, an increase of X by a given percentage, always leads to the same (constant) change of Y. What type of model should you estimate? (2 points)

13a: None of the following answers are correct.

13b: $\log(Y) = \beta_0 + \beta_1 X + u$

13c: $Y = \beta_0 + \beta_1 X + u$

13d: $Y = \exp(\beta_0 + \beta_1 X + u)$

13e: $\exp(Y) = \beta_0 + \beta_1 X + u$

Question 14: You estimate model $M_1 : Y = \beta_0 + \beta_1 X_1 + u$ where X_1 is a continuous variable. You find for model M_1 that $\hat\beta_1 = 1.3$. Now you add another explanatory variable X_2 to model M_1 and estimate $M_2 : Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$. You expect $\beta_2 < 0$. You expect no correlation between X_1 and X_2. What do you expect for model M_2? (3 points)

	a	b	c	d	e
14:	other value	$\hat\beta_1 > 1.3$	$\hat\beta_1 < 1.3$	$\hat\beta_1 = 1.3$	$\hat\beta_1 > \beta_2$

Question 15: You estimate the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$. You obtain the following output from the regression:

```
Call: lm(formula = y ~ x1 + x2)
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 54.8494   11.9916    4.57  0.0013
x1          -13.6813   2.2788   -6.00  0.0002
x2           -6.1135   1.2678   -4.82  0.0009
Residual standard error: 14.92 on 9 degrees of freedom
Multiple R-squared: 0.9115, Adjusted R-squared: 0.8919
F-statistic: 46.37 on 2 and 9 DF, p-value: 1.82e-05
```

You obtain the following output from the hypothesis test:

```
Linear hypothesis test
Hypothesis: x1 - x2 = 0
Model 1: restricted model
Model 2: y ~ x1 + x2
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     10 3434.9
2      9 2003.6  1    1431.4 6.4298 0.03194 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05
                                   '.' 0.1 ' ' 1
```

Which of the following Null hypotheses can be rejected (use a significance level of 5%)? (3 points)

	a	b	c	d	e
15:	other value	$X_1 = X_2$	$\beta_1 = -\beta_2$	$X_1 = -X_2$	$\beta_1 = \beta_2$

11 MW24.1, November 2018 Midterm

Question 1: According to your (two-sided) Null hypothesis the population mean of a normally distributed random variable is `mu0`. In a sample with 15 observations you find a sample standard deviation of 7 and a sample mean of `xBar` which is smaller than `mu0`. When can you reject this Null hypotheses at a level of significance of 10%? (3 points)

1a: None of the following answers are correct.

1b: `mu0 - xBar < qt(.95,df=14)*7/sqrt(15)`

1c: `mu0 - xBar > qt(.95,df=14)*7/sqrt(15)`

1d: `xBar - mu0 < qt(.95,df=14)*7/sqrt(15)`

1e: `xBar - mu0 > qt(.95,df=14)*7/sqrt(15)`

Question 2: How can one obtain a p-value for the above hypothesis test? (2 points)

2a: None of the following answers are correct.

2b: `2*qt((xBar-mu0)/sqrt(15),df=14)`

2c: `2*pt((xBar-mu0)/7*sqrt(15),df=14)`

2d: `pt((xBar-mu0)/7*sqrt(15),df=14)`

2e: `2*dt((xBar-mu0)/sqrt(15),df=14)`

Question 3: A variable `x` contains a sample of a random variable X. Consider the following output from R:

```
        One Sample t-test
data:  x
t = 21.578, df = 14, p-value = 3.832e-12
alternative hypothesis: true mean is not equal to -2
95 percent confidence interval:
 1.954632 2.827546
sample estimates:
mean of x
 2.391089
```

Which of the following statements is correct? (2 points)

3a: None of the following answers are correct.

3b: The probability that $X = -2$ is equal to $3.832 \cdot 10^{-12}$.

3c: The probability that $E(X) = -2$ is equal to $3.832 \cdot 10^{-12}$.

3d: If $E(X) = -2$, then the probability to get a sample like `x` or one that is even more adverse to our Null Hypothesis is equal to $3.832 \cdot 10^{-12}$.

3e: If $X = -2$, then the probability to get a sample like `x` or one that is even more adverse to our Null Hypothesis is equal to $3.832 \cdot 10^{-12}$.

Question 4: You use the `lm` function of R to estimate the relationship $Y_i = \beta_0 + \beta_1 X_i + u_i$. The result of your estimation is stored in `est`. Consider the following output from `confint(est)`:

```
                2.5 %   97.5 %
(Intercept) 2.1552824 5.319727
x           0.6990312 2.300943
```

Based on a level of significance of 5% and with a two sided test, which of the following Null hypotheses can be rejected? (more than one answer possible, 5 points)

4:	a	b	c	d	e
	$\beta_1 \neq 2$	$\beta_1 = 0$	$\beta_1 = 6$	$\beta_0 = 1$	$\beta_1 \neq 3$

Question 5: With a different sample, you estimate again $Y_i = \beta_0 + \beta_1 X_i + u_i$. As before, you store your estimation result in `est`. Consider the following output from a linear hypothesis test:

```
Hypothesis: x = 13
Model 1: restricted model
Model 2: y ~ x
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     11 325598
2     10 231045  1     94553 4.0924 0.07063 .
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05
                                       '.' 0.1 ' ' 1
```

Your significance level is 10%. Which of the following statements is correct? (3 points)

5a: None of the following answers are correct.

5b: The probability that the Null hypothesis is correct is 0.07063.

5c: The probability that the alternative hypothesis is correct is 0.07063.

5d: The Null hypothesis of this test was that $\bar{X} = 13$.

5e: The Null hypothesis is rejected.

Question 6: In a sample with 16 observations of a normally distributed random variable one finds a sample standard deviation of 4 and a sample mean of 8. How does one determine the width of a 90%-confidence interval around the sample mean? (2 points)

6a: None of the following answers are correct.

6b: `2*pt(.025,df=15)*4/sqrt(16)`

6c: `qt(.05,df=15)`

6d: `-2*qt(.05,df=15)`

6e: `2*pt(.025,df=15)*16`

Question 7: A random variable X is distributed as follows: $\Pr(X = 1) = \theta$, $\Pr(X = 2) = 1 - \theta$ with $0 \leq \theta \leq 1$. Your sample is $\{1, 1\}$. What is the ML estimator for θ? (4 points)

7:	a	b	c	d	e
	other value	0	1/4	1/2	1

Question 8: A random variable follows a Poisson distribution. The density function of the Poisson distribution is $P_\lambda(k) = \lambda^k \cdot \exp(-\lambda)/k!$. Your sample is $\{2, 2, 3\}$. What is the ML estimator for λ? (4 points)

8:	a	b	c	d	e
	other value	2	7/3	3	0

Question 9: Consider the following estimation result:

```
Call: lm(formula = y ~ x1 * x2)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    -5          5      -1.0   0.3466
x1              3          2       2.0   0.0805
x2             12         12       1.0   0.3466
x1:x2           1          2       0.5   0.6305
Residual standard error: 12.23 on 8 degrees of freedom
Multiple R-squared: 0.6324, Adjusted R-squared: 0.4946
F-statistic: 4.588 on 3 and 8 DF, p-value: 0.03771
```

`x2` is a dummy which is 0 for group A and 1 for group B. What is the marginal effect of `x1` for group A? (1 point)

9:	a	b	c	d	e
	other value	4	15	3	3.5

Question 10: What is the marginal effect of `x1` for group B? (1 point)

10:	a	b	c	d	e
	other value	15	3	3.5	4

Question 11: Assume that the same observations are coded differently. `x2` is now 1 for group A and 0 for group B. How large is then the marginal effect of `x1` for group B? (1 point)

11:	a	b	c	d	e
	other value	3	3.5	4	15

Question 12: Consider the following regression model:

$$\log Y = \beta_0 + \beta_1 \log X_1 + \beta_2 \log X_2 + u$$

You expect that if X_1 increases by 1% that then Y increases by 2.5%. Which value should you expect for $\hat{\beta}_1$? (2 points)

	a	b	c	d	e
12:	other value	250	$2.5/X_2$	$2.5/Y$	0.025

Question 13: Consider still the previous model. You estimate $\hat{\beta}_1 = 3$. Which of the following statements is (approximately) true? (3 points)

13a: None of the following answers are correct.

13b: When X_1 increases by 0.01 units, then Y increases by 3%.

13c: When X_1 increases by 1 unit, then Y increases by 3%.

13d: The elastiticy of Y with respect to X_1 is 0.03.

13e: When Y increases by 1%, then X_1 increases by 3%.

Question 14: You estimate model $M_1 : Y = \beta_0 + \beta_1 X_1 + u$ where X_1 is a continuous variable. You find for model M_1 that $\hat{\beta}_1 = 5$. Now you add another explanatory variable X_2 to model M_1 and estimate $M_2 : Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$. You expect $\beta_2 > 0$. You also expect a negative correlation between X_1 and X_2. What do you expect for model M_2? (3 points)

	a	b	c	d	e
14:	other value	$\hat{\beta}_1 < \beta_2$	$\hat{\beta}_1 > 5$	$\hat{\beta}_1 < 5$	$\hat{\beta}_1 = 5$

Question 15: You estimate the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$. You obtain the following output from the regression:

```
Call: lm(formula = y ~ x1 + x2)
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 42.9540    14.2642    3.01  0.0088
x1          -4.5090     2.3633   -1.91  0.0757
x2          -5.8428     1.6611   -3.52  0.0031
Residual standard error: 21.89 on 15 degrees of freedom
Multiple R-squared: 0.7839, Adjusted R-squared: 0.755
F-statistic: 27.2 on 2 and 15 DF, p-value: 1.025e-05
```

You obtain the following output from the hypothesis test:

```
Linear hypothesis test
Hypothesis: x1  + x2 = 0
Model 1: restricted model
Model 2: y   x1 + x2
  Res.Df   RSS Df Sum of Sq      F    Pr(>F)
1     16 26584
2     15  7187  1     19397 40.483 1.275e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05
                                         ‘.’ 0.1 ‘ ’ 1
```

Which of the following Null hypotheses can be rejected (use a significance level of 5%)? (3 points)

	a	b	c	d	e
15:	other value	$\beta_1 = \beta_2$	$X_1 = X_2$	$\beta_1 = -\beta_2$	$X_1 = -X_2$

12 MW24.1, November 2017 Midterm Resit

Question 1: In a sample with 11 observations of a normally distributed random variable you observe a sample standard deviation of 3 and a sample mean of 5. According to your (two-sided) Null hypothesis the population mean is $\mu = 20$. Which of the following statements tells you whether you can reject this Null hypotheses at a level of significance of 5%? (3 points)

1a: None of the following answers are correct.

1b: `5-20 > qt(.975,df=10)*3/sqrt(11)`

1c: `5-20 < qt(.025,df=10)*3/sqrt(11)`

1d: `5-20 > qt(.025,df=10)*3/sqrt(11)`

1e: `5-20 < qt(.975,df=10)*3/sqrt(11)`

Question 2: How can one obtain a p-value for the above hypothesis test? (2 points)

2a: None of the following answers are correct.

2b: `2*pt(0.975,(5-20)/3*sqrt(11),df=10)`

2c: `1-2*qt((5-20)/3*sqrt(11),df=10)`

2d: `2*pt((5-20)/3*sqrt(11),df=10)`

2e: `2*qt((5-20)/3*sqrt(11),df=10)`

Question 3: A variable `x` contains a sample of a random variable X. Consider the following output:

```
        One Sample t-test
data:  x
t = -2.0419, df = 5, p-value = 0.09663
alternative hypothesis: true mean is not equal to 7
95 percent confidence interval:
5.159827 7.210913
sample estimates:
mean of x
6.18537
```

Which of the following statements is correct? (2 points)

3a: None of the following answers are correct.

3b: If $X = 7$, then the probability to get a sample like `x` or one that is more adverse to our Null Hypothesis is 0.09663

3c: The probability that $X = 7$ is 0.09663

3d: The probability that $E(X) = 7$ is 0.09663

3e: If $E(X) = 7$, then the probability to get a sample like `x` or one that is more adverse to our Null Hypothesis is 0.09663

Question 4: In a sample with 7 observations of a normally distributed random variable one finds a sample standard deviation of 2 and a sample mean of 18. How does one determine the lower boundary for a 95%-confidence interval around the sample mean? (2 points)

4a: None of the following answers are correct.

4b: `18+qt(.025,df=6)*sqrt(7)/2`

4c: `18+qt(.025,df=6)*sqrt(18)/2`

4d: `18+qt(.025,df=6)*2`

4e: `18+qt(.025,df=6)*2/sqrt(7)`

Question 5: For a sample with 15 observations of a normally distributed random variable you find a sample mean of 10. Earlier you have determined the width of the 95%-confidence interval for the mean as 5. What is the upper boundary of a 95%-confidence interval? (2 points)

5a: None of the following answers are correct.

5b: `10+qnorm(.975,df=14)*5`

5c: `12.5`

5d: `10+pt(.975,df=14)*5/2/qt(.975,df=14)`

5e: `10+qt(.975,df=14)*5`

Question 6: A continuous random variable X has the following density function:

$$f(X) = \begin{cases} \frac{2\theta X}{(\theta X^2+1)^2} & \text{if } X \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

A sample contains a single observation: $\{2\}$. What is the ML estimator for θ? (4 points)

6:	a other value	b 1/2	c 2	d 0	e 1/4

Question 7: A discrete random variable X has the following probability density function:

$$P(X) = \begin{cases} \theta & \text{if } X = 1 \\ 1-\theta & \text{otherwise} \end{cases}$$

A sample contains three observations: $\{1, 1, 2\}$. What is the ML estimator for θ? (3 points)

7:	a other value	b 1	c 1/3	d 1/2	e 2/3

Question 8: You estimate model $M_1 : Y = \beta_0 + \beta_1 X_1 + u$ where X_1 is a binary variable. You find for model M_1 that $\hat\beta_1 = 3$. Now you add another explanatory variable X_2 to model M_1 and estimate $M_2 : Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$. You expect $\beta_2 < 0$. You also expect a positive correlation between X_1 and X_2. What do you expect for model M_2? (3 points)

8:	a other value	b $\hat\beta_1 > 3$	c $\hat\beta_1 < 3$	d $\hat\beta_1 = 3$	e $\hat\beta_1 < \beta_2$

Question 9: Consider the following estimation result:

```
Call: lm(formula = y ~ x1 * x2)
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    -4         4        -1.0   0.3343
x1             -2         1        -2.0   0.0653
x2              1         2         0.5   0.6248
x1:x2           4         1         4.0   0.0013
Residual standard error: 6.869 on 14 degrees of freedom
Multiple R-squared: 0.8587, Adjusted R-squared: 0.8284
F-statistic: 28.36 on 3 and 14 DF, p-value: 3.305e-06
```

x2 is a dummy which is 0 for group A and 1 for group B. What is the marginal effect of `x1` for group A? (1 point)

9:	a other value	b 1	c 2	d 4	e -2

Question 10: What is the marginal effect of `x1` for group B? (1 point)

10:	a other value	b 2	c 4	d -2	e 1

Question 11: Assume that the same observations are coded differently. `x2` is now 1 for group A and 0 for group B. How large is then the marginal effect of `x1` for group B? (1 point)

11:	a other value	b 4	c -2	d 1	e 2

Question 12: You estimate the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$. Consider the following output from the regression and the hypothesis test:

```
Call: lm(formula = y ~ x1 + x2)
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 22.3656    44.1958    0.51   0.6238
x1          16.4450     7.4040    2.22   0.0506
x2         -13.3824     5.1865   -2.58   0.0274
Residual standard error: 50.27 on 10 degrees of freedom
Multiple R-squared: 0.4189, Adjusted R-squared: 0.3027
F-statistic: 3.605 on 2 and 10 DF, p-value: 0.06625
        Linear hypothesis test
        Hypothesis: x1 - x2 = 0
        Model 1: restricted model
        Model 2: y ~ x1 + x2
          Res.Df  RSS  Df Sum of Sq   F     Pr(>F)
        1   11    42452
        2   10    25269 1    17183   6.7999 0.02615 *
```

Which of the following Null hypotheses can be rejected (use a significance level of 5%)? (3 points)

12:	a other value	b $\beta_1 = \beta_2$	c $X_1 = X_2$	d $\beta_1 = -\beta_2$	e $X_1 = -X_2$

Question 13: Still consider the output from the previous question. Which of the following Null hypotheses can be rejected (use a significance level of 5%)? (more than one answer possible, 5 points)

13a: `x1` has no effect on `y`.

13b: The marginal effect of `x1` on `y` is 8.

13c: The intercept is 22.

13d: `x1` has no effect on `x2`.

13e: `x2` has no effect on `y`.

13 MW24.1, November 2017 Midterm

Question 1: Consider the following output:

```
  One Sample t-test
data: X
t = -0.97581, df = 49, p-value = 0.3339
alternative hypothesis: true mean is not equal to 30
95 percent confidence interval:
27.44659 30.88418
sample estimates:
mean of x
29.16538
```

Which of the following statements are correct? (3 points)

1a: None of the following answers are correct.

1b: The probability that $E(X) \neq 30$ is 0.3339.

1c: If $E(X) = 30$, then the probability to get a sample like `X` or one that is even more adverse to our Null hypothesis is 0.3339.

1d: The Null-hypothesis $E(X) \neq 30$ can be rejected.

1e: The probability that $E(X) = 30$ is 0.3339.

Question 2: You use the `lm` function of R to estimate the relationship $Y_i = \beta_0 + \beta_1 X_i + u_i$. The result of your estimation is stored in `est`. Consider the following output from `confint(est)`:

```
                 2.5 %   97.5 %
(Intercept) 4.371154 8.927345
x           5.153215 6.069226
```

Based on a level of significance of 5% and with a two sided test, which of the following Null hypotheses can be rejected? (more than one answer possible, 5 points)

	a	b	c	d	e
2:	$\beta_1 \neq 6$	$\beta_0 = 4$	$\beta_1 \neq 8$	$\beta_1 = 8$	$\beta_1 = 6$

Question 3: You estimate the linear relationship $Y = \beta_0 + \beta_1 X + u$ where the u are i.i.d. and normally distributed. Your sample contains 16 pairs of Y and X. You estimate $\hat{\beta}_1 = -2$ with a standard deviation $\hat{\sigma}_{\hat{\beta}_1} = 2$. Your Null hypothesis is $\beta_1 = -4$. How do you calculate a p-value for a two-sided test of your Null hypothesis? (3 points)

3a: None of the following answers are correct.

3b: `2*pt(1,df=14)`

3c: `2*pt(-1,df=14)`

3d: `2*pt(2,df=14)`

3e: `2*pt(-2,df=14)`

Question 4: Consider still the previous problem. How do you determine the upper boundary of a confidence interval for β_1 when you confidence level is 95%? (2 points)

4a: None of the following answers are correct.

4b: `-2+qt(.95,df=14)`

4c: `-2+pt(.95,df=14)`

4d: `-2+pt(.975,df=14)`

4e: `-2+qt(.975,df=14)`

Question 5: A random variable follows an Exponential distribution with density function $f_\lambda(x) = \lambda \exp(-\lambda x)$ if $x \geq 0$ and 0 otherwise. Your sample contains three observations: $\{0, 0, 1\}$. What is the maximum likelihood estimator for λ? (4 points)

	a	b	c	d	e
5:	other value	1	3	0	1/3

Question 6: A random variable X follows the distribution $\Pr(A|\theta) = \theta, \Pr(B|\theta) = 2\theta, \Pr(C|\theta) = 1 - 3\theta$ with $\theta \in [0, 1/3]$. You observe n times an A and n times a C. You see no B. What is your ML estimator for θ? (4 points)

	a	b	c	d	e
6:	other value	1/2	0	1/6	1/3

Question 7: You estimate the following regression model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + u$$

X_1 is a dummy variable which is 0 for employed and 1 for unemployed people. X_2 is a dummy variable which is 1 if the person participated in training and 0 otherwise. You assume $\beta_0 = 11, \beta_1 = -5, \beta_2 = 2, \beta_3 = 1$. Which value of Y do you expect for an employed person who did not participate in training? (2 points)

	a	b	c	d	e
7:	other value	6	8	9	11

Question 8: Which value of Y do you expect for an unemployed person who did participate in training? (2 points)

	a	b	c	d	e
8:	other value	8	9	11	6

Question 9: Now you code X_1 differently: $X_1 = 0$ for unemployed and $X_1 = 1$ for employed. What do you assume now for β_1? (2 points)

	a	b	c	d	e
9:	other value	3	5	-5	2

Question 10: What do you assume now for β_2? (2 points)

	a	b	c	d	e
10:	other value	7	-2	2	3

Question 11: Consider the following regression model:

$$\log Y = \beta_0 + \beta_1 X_1 + \beta_2 \log X_2 + u$$

You expect that if X_2 increases by 1% that then Y decreases by 3%. Which value should you expect for $\hat{\beta}_2$? (2 points)

	a	b	c	d	e
11:	other value	-3	0.03	$3/X_2$	$3/Y$

Question 12: Consider still the previous model. You estimate $\hat{\beta}_1 = 2$. Which of the following statements is (approximately) true? (3 points)

12a: None of the following answers are correct.

12b: When Y increases by 1%, then X_1 increases by 2%.

12c: When X_1 increases by 0.01 units, then Y increases by 2%.

12d: When X_1 increases by 1 unit, then Y increases by 2%.

12e: The elasticity of Y with respect to X_1 is 2.

Question 13: You estimate model $M_1 : Y = \beta_0 + \beta_1 X_1 + u$. You find for model M_1 that $\hat{\beta}_1 = 8$ and $\hat{\sigma}_{\hat{\beta}_1} = 1$. Now you add another explanatory variable X_2 and estimate model $M_2 : Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$. You know that $\beta_2 < 0$ and that X_1 and X_2 are positively correlated. What can you say about model M_2? (3 points)

	a	b	c	d	e
13:	other value	$\hat{\beta}_1 = 8$	$\hat{\beta}_1 < \hat{\beta}_2$	$\hat{\beta}_1 > 8$	$\hat{\beta}_1 < 8$

14 MW24.1, November 2016 Midterm Resit

Question 1: In your sample with 8 observations of a normally distributed random variable you observe a sample standard deviation of 5 and a sample mean of 21. According to your (two-sided) Null hypothesis the population mean $\mu = 15$. Which of the following statements tells you whether you can reject your Null hypothesis on a level of significance of 5%?

(3 points)

1a: None of the following answers are correct.

1b: `15 > 21+qt(.975,df=7)*5/sqrt(8)`

1c: `15 > 21+qt(.025,df=7)*5/sqrt(8)`

1d: `15 < 21+qt(.025,df=7)*5/sqrt(8)`

1e: `15 < 21+qt(.975,df=7)*5/sqrt(8)`

Question 2: How do you obtain a p-value for the above hypothesis test? (2 points)

2a: None of the following answers are correct.

2b: `qt((15-21)/(5/sqrt(8)),df=7)`

2c: `1-pt((21-15)/(5*sqrt(8)),df=7)`

2d: `pt((15-21)/(5/sqrt(8)),df=7)`

2e: `1-qt((21-15)/(5*sqrt(8)),df=7)`

Question 3: Consider the following output:

```
    One Sample t-test
data:  X
t = -1.207, df = 55, p-value = 0.2326
alternative hypothesis: true mean is not equal to 30
95 percent confidence interval:
26.31195 30.91543
sample estimates:
mean of x
28.61369
```

Which of the following statements are correct? (3 points)

3a: None of the following answers are correct.

3b: The probability that $E(X) = 30$ is 0.2326

3c: The probability that $E(X) \neq 30$ is 0.2326

3d: If $E(X) = 28.61369$ then the probability to get a sample like `X` or one that is even more adverse to our Null hypothesis is 0.2326.

3e: The Null-hypothesis $E(X) \neq 30$ can be rejected.

Question 4: From a random variable X you take a sample of size 16. In your sample you observe a mean of 3 and a standard deviation of 9. Your Null hypothesis is $E(X) = 5$. Which of the following statements tell you whether you can reject your Null hypothesis on a level of significance of 5%? (3 points)

4a: None of the following answers are correct.

4b: `(3-5)/9 < qt(.025,15)`

4c: `(3-5)/(9/4) < qt(.025,15)`

4d: `(3-5)/4 < qt(.975,15)`

4e: `(3-5)/(9/4) < qt(.975,15)`

Question 5: A random variable X follows the distribution $\Pr(1|\theta) = \theta$, $\Pr(2|\theta) = \theta$, $\Pr(3|\theta) = 1 - 2\theta$. You know that $\theta \in [0, 1/2]$. Your sample contains three observations: $\{1, 2, 3\}$. What is your ML estimator for θ? (4 points)

5:	a other value	b 1/2	c 0	d 1/6	e 1/3

Question 6: A random variable follows a Poisson distribution. The density function of the Poisson distribution is $P_\lambda(k) = \lambda^k \cdot \exp(-\lambda)/k!$. Your sample is $\{1, 2, 2\}$. What is the ML estimator for λ? (4 points)

6:	a other value	b 0	c 4/3	d 5/3	e 5/2

Question 7: Consider the following regression model:

$$\log Y = \beta_0 + \beta_1 X_1 + \beta_2 \log X_2 + u$$

You expect that whenever X_1 increases by one unit, then Y increases by 5%. Which value should you expect for $\hat{\beta}_1$? (2 points)

7:	a other value	b 5	c $\exp 0.05$	d $\log 0.05$	e 0.05

Question 8: You estimate $\hat{\beta}_2 = -2$. Which of the following statements is true? (3 points)

8a: None of the following answers are correct.

8b: When X_2 increases by 1%, then Y increases by 0.02 units.

8c: When X_2 increases by 1 unit, then Y increases by 2%.

8d: The elastiticy of Y with respect to X_2 is -2.

8e: When Y increases by 1%, then X_2 increases by 2%.

Question 9: Consider the following regression output. Your significance level is 5%:

```
Call: lm(formula = y ~ x1 + x2)
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 4.5368    10.0543    0.45  0.6625
x1          2.1483     1.5068    1.43  0.1877
x2          3.2215     1.2919    2.49  0.0342
Residual standard error: 12.39 on 9 degrees of freedom
Multiple R-squared: 0.7584, Adjusted R-squared: 0.7047
F-statistic: 14.12 on 2 and 9 DF, p-value: 0.001676
```

Which of the following Null hypotheses can be rejected? (more than one answer possible, 5 points)

9a: `x1` has no effect on `y`.

9b: The effect of `x1` on `y` is not linear.

9c: `x2` has no effect on `y`.

9d: The marginal effect of `x2` on `y` is 10.

9e: The intercept is 1.

Question 10: You estimate model $M_1 : Y = \beta_0 + \beta_1 X_1 + u$. You find for model M_1 that $\hat{\beta}_1 = -2$ and $\hat{\sigma}_{\hat{\beta}_1} = 2$. Now you add another explanatory variable X_2 and estimate model $M_2 : Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$. You know that $\beta_2 > 0$ and that X_1 and X_2 are positively correlated. What do you expect for model M_2? (3 points)

10:	a other value	b $\hat{\beta}_1 < -2$	c $\hat{\beta}_1 > -2$	d $\hat{\beta}_1 = -2$	e $\hat{\beta}_1 > \hat{\beta}_2$

Question 11: Consider the following regression model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + u$$

X_1 is a dummy variable which is 0 for males and 1 for females. X_2 is a dummy variable which is 0 for employed and 1 for unemployed. You assume $\beta_0 = 10$, $\beta_1 = 5$, $\beta_2 = -2$, $\beta_3 = 1$. Which value of Y do you expect for employed males? (2 points)

11:	a other value	b 10	c 15	d 1	e 8

Question 12: Which value of Y do you expect for unemployed males? (2 points)

12:	a	b	c	d	e
	other value	9	10	7	8

Question 13: Which value of Y do you expect for unemployed females? (2 points)

13:	a	b	c	d	e
	other value	16	13	14	15

---

15 MW24.1, November 2016 Midterm

Question 1: In your sample with 12 observations of a normally distributed random variable with known standard deviation of 6 you find a sample mean of 21. Your confidence level is 95%. How do you determine the lower boundary of a confidence interval for the mean?

(2 points)

1a: None of the following answers are correct.

1b: `21+qnorm(.025)*6/sqrt(12)`

1c: `21+qt(.025,11)*6/sqrt(12)`

1d: `21-qnorm(.025)*sqrt(12)/6`

1e: `21-qt(.95,11)*sqrt(12)/6`

Question 2: You use the `lm` function of R to estimate the relationship $Y_i = \beta_0 + \beta_1 X_i + u_i$. The result of your estimation is stored in `est`. Consider the following output from `confint(est)`:

```
                    2.5 %    97.5 %
(Intercept) -454.767569 368.51925
X              8.463561  16.87696
```

Based on a level of significance of 5% and with a two sided test, which of the following Null hypotheses can be rejected?

(more than one answer possible, 5 points)

2:	a	b	c	d	e
	$E(\beta_1) \neq 6$	$E(\beta_1) = 8$	$E(\beta_1) = 6$	$E(\beta_0) = -20$	$E(\beta_1) \neq 8$

Question 3: With a different sample, you estimate again $Y_i = \beta_0 + \beta_1 X_i + u_i$. As before, you store your estimation in `est`. Consider the following output from a linear hypothesis test:

```
Hypothesis: x = 1.8
Model 1: restricted model
Model 2: y ~x
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1 32 300276
2 31 295506 1    4769.6 0.5004 0.4846
```

Your significance level is 5%. Which of the following statements are correct?

(3 points)

3a: The Null hypothesis of this test was that $\bar{X} = 1.8$.

3b: The probability that the Null hypothesis is correct is 0.4846.

3c: The probability that the alternative hypothesis is correct is 0.5004.

3d: The Null hypothesis of this test was that $\beta_1 = 1.8$.

3e: The Null hypothesis is rejected.

Question 4: You estimate the linear relationship $Y = \beta_0 + \beta_1 X + u$ where the u are i.i.d. and normally distributed. Your sample contains 9 pairs of Y and X. You estimate $\hat{\beta}_1 = 10$ with a standard deviation $\hat{\sigma}_{\hat{\beta}_1} = 3$. Your Null hypothesis is $E(\beta_1) = 8$. How do you calculate a p-value for a two-sided test of your Null hypothesis? (4 points)

4a: None of the following answers are correct.

4b: `pt(-2,df=7)`

4c: `pt(2/3,df=7)`

4d: `pt(-2/3,df=7)`

4e: `pt(2,df=7)`

Question 5: The density function of a random variable X is given as $f(X) = \theta/((\theta^2 x^2 + 1)\,\pi)$. Your sample contains a single observation: $\{4\}$. What is the ML estimator for θ?

(4 points)

5:	a	b	c	d	e
	other value	1/4	1/2	4	16

Question 6: A random variable follows a Poisson distribution. The density function of the Poisson distribution is $P_\lambda(k) = \lambda^k \cdot \exp(-\lambda)/k!$. Your sample is $\{2, 2, 2\}$. What is the ML estimator for λ?

(4 points)

6:	a	b	c	d	e
	other value	1/3	2	5/2	1/2

Question 7: Consider the following regression model:

$$\log Y = \beta_0 + \beta_1 \log X_1 + \beta_2 \log X_2 + u$$

You expect that whenever X_1 increases by one unit, then Y increases by 5%. Which value should you expect for $\hat{\beta}_1$?

(2 points)

7:	a	b	c	d	e
	other value	$\exp 0.05$	$\log 0.05$	0.05	5

You estimate $\hat{\beta}_2 = 6$. Which of the following statements is true?

(3 points)

8a: None of the following answers are correct.

8b: When X_2 increases by 1 unit, then Y increases by 6%.

8c: The elastiticy of Y with respect to X_2 is 6.

8d: When Y increases by 1%, then X_2 increases by 6%.

8e: When X_2 increases by 1%, then Y increases by 0.06 units.

Question 9: Consider the following regression output. Your significance level is 5%:

```
Call: lm(formula = y ~ x1 + x2)
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -4.6189    4.4575    -1.04  0.3245
x1           6.3709    0.7812     8.16  0.0000
x2           2.6126    0.5189     5.03  0.0005
Residual standard error: 7.313 on 10 degrees of freedom
Multiple R-squared: 0.9578, Adjusted R-squared: 0.9493
F-statistic: 113.4 on 2 and 10 DF, p-value: 1.343e-07
```

Which of the following Null hypotheses can be rejected?

(more than one answer possible, 5 points)

9a: x1 has no effect on y.

9b: x2 has no effect on y.

9c: The marginal effect of x1 on y is -1.

9d: The intercept is -1.

9e: The effect of x1 on y is not linear.

Question 10: Consider the following regression model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + u$$

X_1 and X_2 are dummy variables which are either 0 or 1. You observe the following means for Y:

	$X_2 = 0$	$X_2 = 1$
$X_1 = 0$	$\bar{Y} = 4$	$\bar{Y} = 3$
$X_1 = 1$	$\bar{Y} = 2$	$\bar{Y} = 5$

What is the value of $\hat{\beta}_0$? (2 points)

	a	b	c	d	e
10:	other value	-1	2	4	-2

Question 11: What is the value of $\hat{\beta}_1$? (2 points)

	a	b	c	d	e
11:	other value	2	4	-2	-1

Question 12: What is the value of $\hat{\beta}_2$? (2 points)

	a	b	c	d	e
12:	other value	4	-2	-1	2

Question 13: What is the value of $\hat{\beta}_3$? (3 points)

	a	b	c	d	e
13:	other value	-2	-1	2	4

---

16 MW24.1, November 2015 Midterm Resit

Question 1: In your sample with 16 observations of a normally distributed random variable you find a mean of 63 and a standard deviation of 9. Your confidence level is 99%. How do you determine the width of a confidence interval for the mean? (2 points)

1a: None of the following answers are correct.

1b: `qt(.995,15)*18`

1c: `qnorm(.995)*18`

1d: `qnorm(.9)*32/3`

1e: `qt(.995,15)*18/4`

Question 2: Use a significance level of 10%. Your variable x contains a sample of a random variable X. Consider the following output:

```
    One Sample t-test
data: x
t = 0.10885, df = 99, p-value = 0.9135
alternative hypothesis: true mean is not equal to 47
90 percent confidence interval:
46.31434 47.78186
sample estimates:
mean of x
47.0481
```

Which of the following statements are correct:

(more than one answer possible, 5 points)

2a: The Null hypothesis $E(X) = 47$ can be rejected.

2b: The probability to get a sample like x or one that is even more adverse to our Null hypothesis is 0.9135.

2c: If $E(X) = 47$ then the probability to get a sample like x or one that is even more adverse to our Null hypothesis is 0.9135.

2d: The probability that $E(X) = 47$ is 0.9135.

2e: The probability that $E(X) \neq 47$ is 0.9135.

Question 3: Still consider the output from the previous question. Use a significance level of 10%. Using a two sided test, which of the following Null hypotheses can be rejected.

(more than one answer possible, 5 points)

	a	b	c	d	e
3:	$E(X) \neq 48$	$E(X) \neq 49$	$E(X) = 48$	$E(X) = 47$	$E(X) = 49$

Question 4: From a random variable X you take a sample of size 16. In your sample you observe a mean of 8 and a standard deviation of 9. Your Null hypothesis is $E(X) = 12$. How do you calculate a p-value for a two-sided test of your Null hypothesis? (4 points)

4a: None of the following answers are correct.

4b: `2*pt(-4/9,df=15)`

4c: `2*pt(-16/9,df=15)`

4d: `2*pt(-4/4,df=15)`

4e: `pt(-4/9,df=15)`

Question 5: A random variable follows an Exponential distribution. The density function of the Exponential distribution is $f_\lambda(x) = \lambda \exp(-\lambda x)$ if $x \geq 0$ and 0 otherwise. Your sample contains a single observation: $\{1\}$. What is the maximum likelihood estimator for λ?

(4 points)

	a	b	c	d	e
5:	other value	1	e	∞	$1/e$

Question 6: The first moment of the Exponential distribution is $\mu = 1/\lambda$. Your sample now contains two observations: $\{2, 3\}$. What is the methods of moments estimator for λ based on the first moment?

(2 points)

	a	b	c	d	e
6:	other value	2	5/2	1	2/5

Question 7: Consider the following regression model:

$$\log Y = \beta_0 + \beta_1 \log X_1 + \beta_2 \log X_2 + u$$

with $X_1 > 0$ and $X_2 > 0$. Which of the following statements are true? (more than one answer possible, 10 points)

7a: The marginal effect of X_1 on Y is β_1.

7b: The marginal effect of X_1 on Y depends on β_0.

7c: The marginal effect of X_1 on Y is always positive.

7d: The marginal effect of X_1 on Y is only positive if β_1 is positive.

7e: The marginal effect of X_1 on Y depends on X_2.

Question 8: You estimate model $M_1 : Y = \beta_0 + \beta_1 X_1 + u$. You find for model M_1 that $\hat{\beta}_1 = 5$ and $\hat{\sigma}_{\hat{\beta}_1} = 1$. Now you add another explanatory variable X_2 and estimate model $M_2 : Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u$. You know that $\beta_2 > 0$ and that X_1 and X_2 are negatively correlated. What do you expect for model M_2?

(3 points)

	a	b	c	d	e
8:	other value	$\hat{\beta}_1 > 5$	$\hat{\beta}_1 < 5$	$\hat{\beta}_1 = 5$	$\hat{\beta}_1 > \hat{\beta}_2$

Question 9: Now assume that $\beta_2 = 0$. What do you expect for model M_2?

(3 points)

	a	b	c	d	e
9:	other value	$\hat{\beta}_1 < 5$	$\hat{\beta}_1 = 5$	$\hat{\beta}_1 > \hat{\beta}_2$	$\hat{\beta}_1 > 5$

Question 10: Still $\beta_2 = 0$. What do you expect for $\hat{\sigma}_{\hat{\beta}_1}$ from model M_2?

(3 points)

	a	b	c	d	e
10:	other value	< 1	indeterminate	$= 1$	> 1

Question 11: Consider the following estimation result:

```
Call: lm(formula = y ~ x1 * x2)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -1         3      -0.333  0.7403
x1            12         1      12.000  0.0000
x2             7         7       1.000  0.3223
x1:x2         -3         1      -3.000  0.0043
Residual standard error: 9.64 on 48 degrees of freedom
Multiple R-squared: 0.9073, Adjusted R-squared: 0.9015
F-statistic: 156.6 on 3 and 48 DF, p-value: < 2.2e-16
```

`x2` is a dummy which is 0 for group A and 1 for group B. What is the marginal effect of `x1` for group A? (1 point)

	a	b	c	d	e
11:	other value	12	-3	9	11

Question 12: What is the marginal effect of `x1` for group B?

(2 points)

	a	b	c	d	e
12:	other value	-3	9	11	12

Question 13: Now you code your data differently. `x2` is now 1 for group A and 0 for group B. How large is now the marginal effect of `x1` for group B? (2 points)

	a	b	c	d	e
13:	other value	9	12	15	-7

17 MW24.1, November 2015 Midterm

Question 1: In your sample with 9 observations of a normally distributed random variable with known standard deviation of 16 you determine a mean of 36. Your confidence level is 90%. How do you determine the width of a confidence interval for the mean?

(2 points)

1a: None of the following answers are correct.

1b: `qnorm(.95)*32`

1c: `qt(.9)*32`

1d: `qnorm(.95)*32/3`

1e: `qt(.95)*32`

Question 2: Use a significance level of 5%. Your variable `x` contains a sample of a random variable X. Consider the following output:

```
        One Sample t-test
data:  x
t = -2.2961, df = 99, p-value = 0.02378
alternative hypothesis: true mean is not equal to 100
95 percent confidence interval:
96.09283 99.71528
sample estimates:
mean of x
97.90406
```

Which of the following statements are correct: (more than one answer possible, 5 points)

2a: The Null hypothesis $E(X) = 100$ can be rejected.

2b: If $E(X) = 100$ then the probability to get a sample like `x` or one that is even more adverse to our Null Hypothesis is 0.02378.

2c: The probability that $E(X) = 100$ is 0.02378.

2d: The probability that $E(X) \neq 100$ is 0.02378.

2e: The probability to get a sample like `x` or one that is even more adverse to our Null Hypothesis is 0.02378.

Question 3: Still consider the output from the previous question. Use a significance level of 5%. Using a two sided test, which of the following Null hypotheses can be rejected.

(more than one answer possible, 5 points)

	a	b	c	d	e
3:	$E(X) \neq 95$	$E(X) = 96$	$E(X) = 97$	$E(X) = 99$	$E(X) \neq 101$

Question 4: From a random variable X you take a sample of size 9. In your sample you observe a mean of 5 and a standard deviation of 4. Your Null hypothesis is $E(X) = 10$. How do you calculate a p-value for a two-sided test of your Null hypothesis?

(4 points)

4a: None of the following answers are correct.

4b: `2*pt(-15/4,df=8)`

4c: `2*pt(-5/3,df=8)`

4d: `pt(-15/4,df=8)`

4e: `2*pt(-5/4,df=8)`

Question 5: A random variable X is distributed as follows: $\Pr(X=1)=\theta, \Pr(X=2)=1-\theta$. Your sample is $\{1,2,2,2\}$. What is the ML estimator for θ?

(4 points)

5:	a other value	b 1/2	c 3/4	d 0	e 1/4

Question 6: A random variable follows a Poisson distribution. The density function of the Poisson distribution is $P_\lambda(k)=\lambda^k\cdot\exp(-\lambda)/k!$. Your sample is $2,3$. What is the ML estimator for λ?

(4 points)

6:	a other value	b 3	c 0	d 2	e 5/2

Question 7: Consider the following regression model:

$$Y=\beta_0+\beta_1\log X_1+\beta_2\log X_2+u$$

You expect that for $X_1=5$ the marginal effect of X_1 on Y is 4. What do you expect for $\hat\beta_1$?

(2 points)

7:	a other value	b 1/4	c 4	d 5	e 20

Question 8: You estimate $\hat\beta_2=6$. What is the marginal effect of X_2 on Y if $X_2=2$?

(2 points)

8:	a other value	b 2	c 3	d 6	e 0

Question 9: Consider the following regression output. Your significance level is 5%:

```
Call: lm(formula = y ~x1 + x2)
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.5466    6.5827     1.60  0.1532
x1           5.1249    2.1124     2.43  0.0457
x2           2.0349    1.0665     1.91  0.0980
Residual standard error: 10.32 on 7 degrees of freedom
Multiple R-squared:  0.8978, Adjusted R-squared:  0.8686
F-statistic: 30.74 on 2 and 7 DF, p-value: 0.0003415
```

Which of the following statements are correct:

(more than one answer possible, 5 points)

9a: `x1` has a significant effect on `y`.

9b: The hypothesis, the marginal effekt of `x2` on `y` is 1, can be rejected.

9c: The hypothesis, the effect of `x1` on `y` was not linear can be rejected.

9d: `x2` has a significant effect on `y`.

9e: The hypothesis, the marginal effect of `x1` on `y` is -5, can be rejected.

Question 10: Consider the following regression model:

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_1X_2+u$$

X_1 and X_2 are dummy variables which are either 0 or 1. You observe the following means for Y:

X_1	X_2	$\bar Y$
0	0	2
1	0	5
0	1	2
1	1	5

What is the value of $\hat\beta_0$? (2 points)

10:	a other value	b 3	c 0	d 1	e 2

Question 11: What is the value of $\hat\beta_1$? (2 points)

11:	a other value	b 0	c 1	d 2	e 3

Question 12: What is the value of $\hat\beta_2$? (2 points)

12:	a other value	b 1	c 2	d 3	e 0

Question 13: What is the value of $\hat\beta_3$? (3 points)

13:	a other value	b 0	c 1	d -2	e -1