Name _____

Spring 2017 **STAT 401** **Final exam**
(100 points)

Instructions:

- Full credit will be given only if you show your work.
- The questions are not necessarily ordered from easiest to hardest.
- You are allowed to use any resource except aid from another individual.
- Aid from another individual will automatically earn you a 0.
- Feel free to tear off the last page. There is no need to turn it in.

One-way ANOVA

Suppose you fit two regression models: an intercept-only model and a model with a categorical variable named "Var". The table below provides an estimate for the error variance and its degrees of freedom.

Model	df	$\hat{\sigma}$
Intercept-only	20	3
Intercept with Var	14	2

Use this information to answer the following questions.

1. How many levels of the categorical variable "Var" are there? (1 pts)

2. How many total observations are there? (1 pts)

3. If the design is balanced, how many replicates are there for each level of the categorical variable "Var"? (2 pts)

4. Fill out this one-way ANOVA table below (12 pts)

	SS	df	MS	F	p
Var					
Error					
Total					

5. Interpret this p-value. (4 pts)

Regression diagnostics

The file `diagnostics.csv` contains a set of 5 response variables (`y1`, `y2`, `y3`, `y4`, and `y5`) and a common explanatory variable x. Consider simple linear regression models for each of the five response variables separately. One of the five response variables meets all model assumptions while each of the other four violates exactly one model assumption. For each response, 1) identify the model assumption violation (if any) and 2) describe how you know that assumption is violated, e.g. what diagnostic plot is informative and what does it look like. (4 pts each)

`y1`

`y2`

`y3`

`y4`

`y5`

Wool

For the following questions, please refer to the "Wool - R Code" page. If you need any background information, please see `?warpbreaks` in `R`.

Write down the model that was used in this analysis. Make sure to define any notation you introduce. (20 pts)

Wool (continued)

Provide an interpretation for the following quantities (4 pts each):

- 39.278

- 11.62

- (-17.77790,-2.2221006)

- 26.38889

- 4.722222

Donation

For the following questions, please use the `donation.csv` data file. These data are filtered version of the data used in the Data Mining competition (see `https://archive.ics.uci.edu/ml/machine-learning-databases/kddcup98-mld/kddcup98.html`) where the filtering only includes Iowa. In addition, only three variables remain: the donation amount from the last targetted mailing (TARGET_D), the type of neighborhood the donor lives in (DOMAIN), and a measure of the donor's wealth (WEALTH2). Fit a linear regression model using log(TARGET_D+1) as the response and DOMAIN and WEALTH2 as the explanatory variables treating WEALTH2 as continuous.

1. Write the R code you used to fit this model. (4 pts)

2. Provide an estimate for the multiplicative effect of a one-unit increase in WEALTH2 level on the median TARGET_D. (4 pts)

3. Provide a 95% credible interval for a contrast estimate to compare mean log(TARGET_D+1) for rural (R1,R2, and R3) vs city (C1, C2, and C3) domains averaged over WEALTH2. (8 pts)

4. Why might this model not be appropriate for these data? (4 pts)

(intentionally left blank)

Wool - R Code

```
library("emmeans")
```

```
## Error in library("emmeans"):  there is no package called 'emmeans'
```

```
m <- lm(breaks ~ wool + tension, data = warpbreaks)
summary(m)
##
## Call:
## lm(formula = breaks ~ wool + tension, data = warpbreaks)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -19.500  -8.083  -2.139   6.472  30.722
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   39.278      3.162  12.423  < 2e-16 ***
## woolB         -5.778      3.162  -1.827 0.073614 .
## tensionM     -10.000      3.872  -2.582 0.012787 *
## tensionH     -14.722      3.872  -3.802 0.000391 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.62 on 50 degrees of freedom
## Multiple R-squared:  0.2691,Adjusted R-squared:  0.2253
## F-statistic: 6.138 on 3 and 50 DF,  p-value: 0.00123
confint(m)
##                 2.5 %     97.5 %
## (Intercept)  32.92715 45.6284061
## woolB       -12.12841  0.5728505
## tensionM    -17.77790 -2.2221006
## tensionH    -22.50012 -6.9443228
(em <- emmeans(m, ~tension))
```

```
## Error in emmeans(m, ~tension):  could not find function "emmeans"
```

```
co <- contrast(em, "pairwise")
```

```
## Error in contrast(em, "pairwise"):  could not find function "contrast"
```

```
confint(co)
```

```
## Error in confint(co):  object 'co' not found
```