

ST512

Fall Semester, 2007

Quiz 1

Name: _____

Directions:

- Answer questions as directed. Please show work, but not when the answer may be found directly in output. Partial credit may be awarded for correct expressions given in incomplete answers.
- Use the back of the page or the extra paper provided if necessary.
- By default, use level of significance $\alpha = 0.05$ on all testing or confidence interval problems, unless directed otherwise.
- If time is an issue, you do not need to complete elementary arithmetic, so that, for example, a confidence interval of the form $10.2 \pm 2.1\sqrt{\frac{2}{5}}11.4$ is an acceptable answer.
- Some problems here are based on experiments discussed in Yandell (1997) and others are taken from the R software package examples.

1. (35 points) The number of eggs (y) produced by a random sample of $n = 10$ gravid female iguanas are observed along with the mother's weight (x) in *kg*. Summary statistics are given below:

$$\bar{x} = 1.375, \quad \sum_i (x_i - \bar{x})^2 = 0.84, \quad \bar{y} = 44.7, \quad \sum_i (y_i - \bar{y})^2 = 968.1, \quad r_{xy} = 0.935$$

- (a) Use Fisher's transformation to test (use level $\alpha = .05$) the hypothesis that weights and egg production are uncorrelated.

- (b) Consider the slope in a simple linear regression of y on x .

i. Report $\hat{\beta}$, the least squares estimate of β .

ii. Give the units of $\hat{\beta}$.

iii. Report an estimate of the standard error of $\hat{\beta}$.

(c) Consider an individual gravid iguana with weight $x = 1$ *kg*. Obtain a 95% prediction interval for the number of eggs she will produce.

(d) The $i = 4^{th}$ mother iguana weighed one *kg* and produced 33 eggs. Give the fitted value and residual for this observation.

(e) Specify the assumptions about E_i in the regression model $y_i = \beta_0 + \beta_1 x_i + E_i$. Indicate which assumption is violated (strictly speaking).

2. On each of $n = 31$ runners randomly sampled from a population of interest, the response $\text{RunTime}(y)$ was measured along with two potential predictors, $\text{RestPulse}(x_1)$ and $\text{O2Uptake}(x_2)$. Consider several linear regression models for mean RunTime :

$$\begin{aligned}
 M_0 \quad \mu(x_1, x_2) &= \beta_0 \\
 M_1 \quad \mu(x_1, x_2) &= \beta_0 + \beta_1 \text{RestPulse} \\
 M_2 \quad \mu(x_1, x_2) &= \beta_0 + \beta_2 \text{O2Uptake} \\
 M_3 \quad \mu(x_1, x_2) &= \beta_0 + \beta_1 \text{RestPulse} + \beta_2 \text{O2Uptake}
 \end{aligned}$$

In the SAS output labelled “RUNNER OUTPUT” at the end of this problem, PROC CORR and PROC REG were used to get correlations and sums of squares.

- (a) (24 pts) For each model comparison below, report an F -ratio and associated degrees of freedom (df_1, df_2). Some calculations may be required.

Comparison	F -ratio	df_1, df_2
M_0 vs M_3		
M_0 vs M_1		
M_1 vs M_3		
M_2 vs M_3		

- (b) (18 pts) Estimate the mean and standard deviation of the population from which these men were sampled using the four models. For those models involving predictor variables, you may assume fixed values of `restpulse= 60` and/or `O2Uptake= 50`.

Model used	Estimated Mean	Estimated Std dev.
M_0		
M_1	11.1	
M_2	10	
M_3		

- (c) (7 pts) Under model M_3 , consider the effect of a *1bpm* decrease in `restpulse` and an increase of 1 units in `O2Uptake` on the mean. Report an estimate of this effect along with a standard error.

RUNNER OUTPUT

```
proc corr data=runners;
  with RunTime;
proc reg data=runners;
  model RunTime = RestPulse O2Uptake /ss1 ss2 covb;
```

The SAS System
The CORR Procedure

1

1 With Variables: RunTime
2 Variables: RestPulse O2Uptake

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
RunTime	31	10.58613	1.38741	328.17000	8.17000	14.03000
RestPulse	31	53.45161	7.61944	1657	40.00000	70.00000
O2Uptake	31	47.37581	5.32723	1469	37.38800	60.05500

Pearson Correlation Coefficients, N = 31
Prob > |r| under H0: Rho=0

	Rest Pulse	O2Uptake
RunTime	0.45038 0.0110	-0.86219 <.0001

The SAS System
The REG Procedure

2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	43.70121	21.85060	43.56	<.0001
Error	28	14.04633	0.50165		
Corrected Total	30	57.74754			

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t	Type I SS	Type II SS
Intercept	1	19.37438	1.88680	10.27	<.0001	3474.04996	52.89418
RestPulse	1	0.02298	0.01851	1.24	0.2248	11.71377	0.77284
O2Uptake	1	-0.21142	0.02648	-7.99	<.0001	31.98744	31.98744

Covariance of Estimates

Variable	Intercept	RestPulse	O2Uptake
Intercept	3.560021852	-0.027590139	-0.043674233
RestPulse	-0.027590139	0.0003426823	0.0001957374
O2Uptake	-0.043674233	0.0001957374	0.0007010277

3. (12 points) In simple linear regression, suppose the predictor variable is centered, so that $x_i^{\text{new}} = x_i - \bar{x}$. Indicate whether each statement below is true or false by writing a T or an F next to each.

- (a) The least squares estimate of the slope is unchanged ($\hat{\beta}_1^{\text{new}} = \hat{\beta}_1$).
- (b) The least squares estimate of the intercept is changed to ($\hat{\beta}_0^{\text{new}} = \bar{y}$).
- (c) The interpretation of the intercept changes.
- (d) The new estimators of the regression coefficients are uncorrelated.
- (e) The error mean square is likely to be reduced.
- (f) The covariance matrix of $\hat{\beta}^{\text{new}}$ is diagonal.

4. (4 points) A response variable has sample mean $\bar{y} = 10$, standard deviation $s_y = 2$ and a sample correlation coefficient with a predictor x of $r_{yx} = 0.6$. Use simple linear regression to estimate the mean at a value $x = x_0$ that is one sample standard deviation (s_x) above the sample mean (\bar{x}) of the observed predictor values.