TEST #1

STA 4853 **Name:**______________________________________

March 4, 2015

Please read the following directions.
DO NOT TURN THE PAGE UNTIL INSTRUCTED TO DO SO

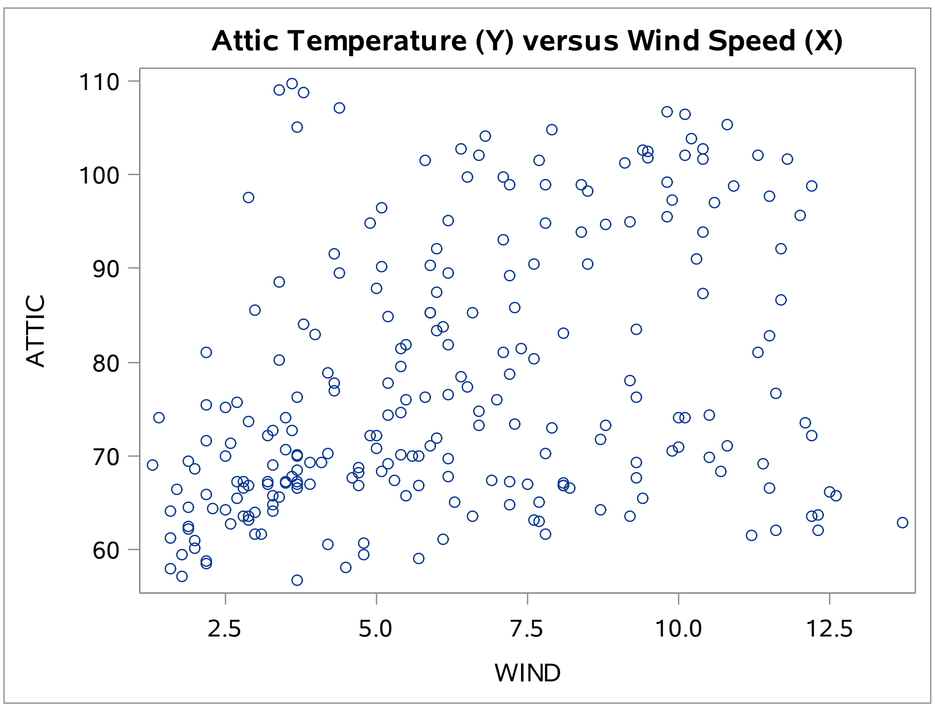
Directions

- This exam is **closed book** and **closed notes**.
- There are 31 questions.
- Circle a **single** answer for each multiple choice question. Your choice should be made clearly.
- On the multiple choice questions, always **circle the correct response**. (Sometimes the question has an empty blank or a box, but this is **NOT** where the answer goes.)
- There is no penalty for guessing.
- The exam has **13** pages.
- Each question is worth equal credit.

Problem 1. Attic temperature and wind speed have been observed at hourly intervals for 240 consecutive hours. The graph given below is a scatter plot of the attic temperature (Y) versus the wind speed (X). What would you expect **solely on the basis of this scatter plot?**

a) The PACF of Y will cutoff to zero.
b) The PACF of X will cutoff to zero.
c)$\star$ In a regression of Y on X, the coefficient of X will be small and positive.
d) In a regression of Y on X, the coefficient of X will be large and negative.
e) The ACF of Y will decay to zero slowly.
f) The ACF of Y will decay to zero rapidly.
g) The time series plot of X will exhibit a strong daily (24 hour) pattern.
h) The ACF of X will decay to zero slowly.
i) The ACF of X will decay to zero rapidly.

1



Problem 2. The variance of a stationary AR(1) process is equal to __________

a) $\dfrac{C}{1-\phi_1}$ **b)**$\star$ $\dfrac{\sigma^2}{1-\phi_1^2}$ **c)** $1+2\sum_{j=1}^{k-1} r_j^2$ **d)** $\dfrac{\sigma}{1-\phi_1}$

e) $\dfrac{C^2}{1-\phi_1^2}$ **f)** $(1+2\sum_{j=1}^{k-1} r_j^2)^{1/2}$ **g)** $n(n+2)\sum_{k=1}^{m} \dfrac{r_k}{n-k}$ **h)** $n\sum_{k=1}^{m} \dfrac{r_k^2}{n-k}$

Problem 3. Suppose that z_t is an AR(1) process generated by

$$z_t = a_t + \phi_1 z_{t-1}\,.$$

By making an appropriate substitution for z_{t-1} one may derive _________

a) $z_t = a_t - \theta_1 a_{t-1}$
b) $z_t = a_t - \theta_1 z_{t-1}$
c) $z_t = a_t - \phi_1 z_{t-1}$
d) $z_t = a_t - \theta_1 a_{t-1} - \theta_1^2 a_{t-2}$
e)$\star$ $z_t = a_t + \phi_1 a_{t-1} + \phi_1^2 z_{t-2}$
f) $z_t = a_t + \phi_1 a_{t-1} + \phi_1 z_{t-2}$
g) $z_t = a_t + \phi_1 z_{t-1} + \phi_1^2 z_{t-2}$
h) $z_t = \phi_1 a_t + \phi_1^2 z_{t-1} + \phi_1^3 z_{t-2}$
i) $z_t = a_t - \theta_1 a_{t-1} - \theta_1^2 a_{t-2} - \theta_1^3 a_{t-3}$

Problem 4. Suppose you have used SAS to estimate the coefficients C, θ_1, θ_2, ϕ_1, ϕ_2 in an ARMA(2,2) model. If the p-value for the estimate of ϕ_2 is large, but the p-values for all the other parameter estimates are small, then we usually _________

a) drop θ_2 from the model
b) drop ϕ_1 from the model
c) add another AR term to the model
d) add another MA term to the model
e) try differencing the series
f)$\star$ drop ϕ_2 from the model
g) retain ϕ_2 in the model
h) try integrating the series
i) try transforming the series

Problem 5. In regression, large values of H (the leverage) identify _________

a) cases with unusual response values
b) influential cases in the data
c)$\star$ cases with unusual covariate values
d) influential covariates in the model
e) covariates which can be dropped
f) covariates with serial correlation
g) covariates which should be retained
h) serial correlation in the residuals

Problem 6. For the time series z_t, the scatter plot of z_{t-10} versus z_t is given below. This is the plot of "z lagged by 10" versus z. From this plot we conclude that ________

a) the series z_t tends to increase with time t

b) the series z_t tends to decrease with time t

c) the residuals are **not** normally distributed

d) the residuals **are** normally distributed

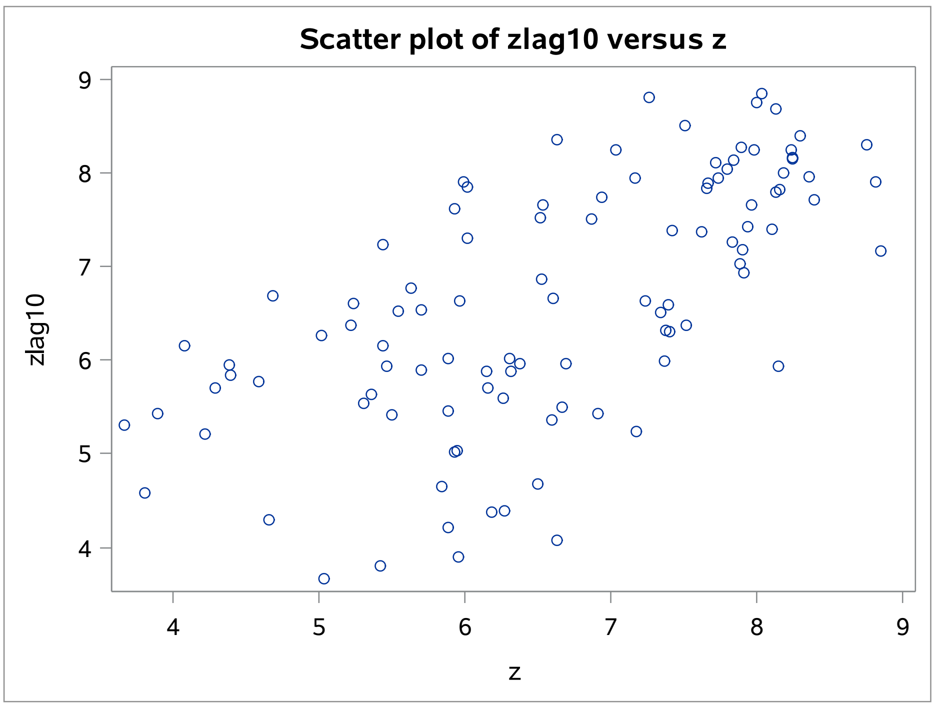
e) the partial autocorrelation at lag 10 is is positive

f) the partial autocorrelation at lag 10 is is negative

g)$\star$ the autocorrelation at lag 10 is positive

h) the autocorrelation at lag 10 is negative

1



Problem 7. Suppose one is given the values of $\rho_1, \rho_2, \ldots, \rho_k$ (the autocorrelations at lags 1 to k) for some stationary process z_t. Then using these values it is possible to compute ________

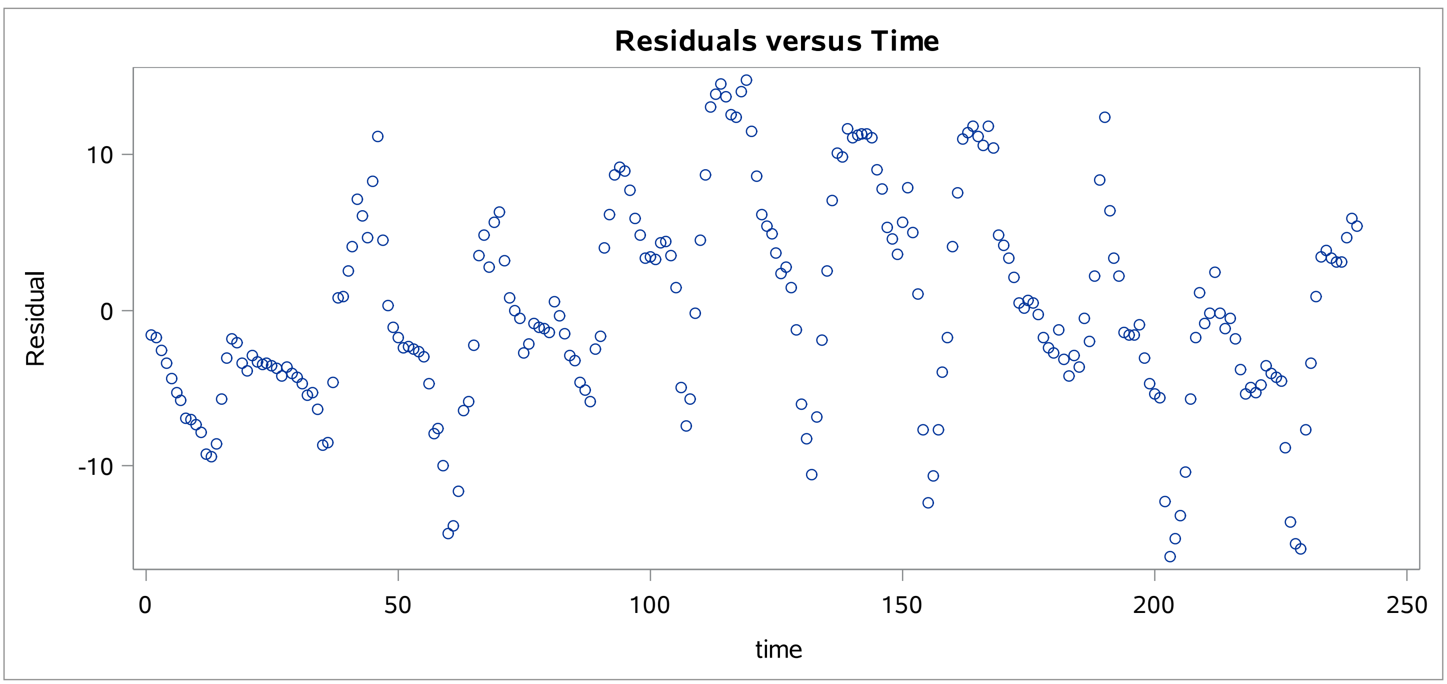
a) μ_z **b)** σ_z^2 **c)** C **d)** σ_a^2 **e)** ρ_{k+1} **f)** r_{k+1} **g)**$\star$ ϕ_{kk} **h)** γ_{k+1}

Problem 8. A response variable Y_t and two covariates $X_{1,t}$ and $X_{2,t}$ are observed at times $t = 1, 2, \ldots, 240$. The regression model

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \varepsilon_t$$

was fit to the data using OLS (ordinary least squares). A plot of the residuals versus time is given below. What would you expect on the basis of this plot?

- **a)** The p-values for $\hat{\beta}_1$ and $\hat{\beta}_2$ will **not** be statistically significant.
- **b)** The p-values for $\hat{\beta}_1$ and $\hat{\beta}_2$ **will** be statistically significant.
- **c)** The response variable Y_t is not stationary.
- **d)** The covariates $X_{1,t}$ and $X_{2,t}$ are not stationary.
- **e)** The residuals **are** normally distributed.
- **f)** The residuals are **not** normally distributed.
- **g)** A log transform of the response variable Y_t is necessary.
- **h)**$\star$ The Durbin-Watson test will detect serial correlation.



Problem 9. If you fit a regression model, SAS will produce a table with a row for each regression parameter β_i in the model. The columns report the parameter estimates $\hat{\beta}_i$, standard errors $\mathrm{SE}(\hat{\beta}_i)$, t-values t_i, and P-values, respectively. In this table, a **large** value of $|t_i|$ will lead to a __________

a) small parameter estimate $\hat{\beta}_i$ **b)** large standard error $\mathrm{SE}(\hat{\beta}_i)$ **c)** small Root MSE

d) large R-squared **e)**$\star$ small P-value **f)** large P-value

Problem 10. Any stationary ARMA(p,q) process can be written as an MA(∞) process

$$z_t = \mu_z + a_t + \sum_{i=1}^{\infty} \psi_i a_{t-i}$$

where _________

a) there is a cutoff in the ACF of z_t after lag q

b) there is a cutoff in the PACF of z_t after lag p

c) $a_t \to \infty$ as $i \to \infty$

d) $\text{Var}(z_t) = \sigma_a^2$

e) The ACF and PACF of z_t is exactly zero for all nonzero lags

f)$\star$ $\psi_i \to 0$ as $i \to \infty$

g) $\psi_i \to \infty$ as $i \to \infty$

h) $\psi_i = -\phi_i$ for $i = 1, 2, 3, \ldots$

Problem 11. Suppose you have a time series $z_1, z_2, \ldots, z_n$ and use SAS to fit an ARMA(1, 1) model to this data and to calculate the corresponding residuals. The residuals you obtain are estimates of the _________

a) ACF   b) constant C c) AR parameters   d) MA parameters

e)$\star$ random shocks a_t f) PACF   g) σ_a^2 h) σ_z^2

Problem 12. If you integrate a stationary ARMA process, the result will be a _________ process.

a) stationary ARMA

b) random shock

c) white noise

d)$\star$ non-stationary AR**I**MA

e) stationary AR**I**MA

f) exponentially decaying

g) alternating exponential

h) exponential growth

Problem 13. Suppose you fit an AR(2) model to some data and you read in the SAS output that the estimated value of the parameter ϕ_2 is $\hat{\phi}_2 = 0.6$, the t-value is 3.0, and the P-value is 0.004. What is the value of the standard error SE($\hat{\phi}_2$)?

a) $\sqrt{3.0}$ b) $\sqrt{0.004}$ c) 0.012   d) 0.00667   e) 5.0   f)$\star$ 0.2   g) 1.8

Problem 14. Suppose you are analyzing a time series $z_1, z_2, \ldots, z_n$ using SAS PROC ARIMA. In the output produced by the IDENTIFY statement, the values of the Ljung-Box test statistics $Q(6)$, $Q(12)$, $Q(18)$, $Q(24)$ are given along with their corresponding P-values. If $\{z_t\}$ is actually just a random shock sequence, then we expect that _________

a)$\star$ all four P-values will be large
b) all four P-values will be small
c) $Q(6)$, $Q(12)$, $Q(18)$, $Q(24)$ will all be large
d) most of the ACF values will lie outside the band
e) most of the PACF values will lie outside the band
f) $H_0 : \rho_1 = \rho_2 = \cdots = \rho_6 = 0$ will be rejected
g) $H_0 : Q(6) = Q(12) = Q(18) = Q(24) = 0$ will be rejected

Problem 15. A realization from a process with a non-stationary mean will usually have a sample ACF _________

a) with most of its values inside the band
b) with all of its values inside the band
c) with a cutoff to zero after some lag
d)$\star$ which decays very slowly to zero
e) which decays very rapidly to zero
f) whose variance changes with the level of the series
g) with alternating exponential decay
h) with sinusoidal decay to zero

Problem 16. Let ρ denote the population correlation between X and Y. If $\rho = 0$, then we know (roughly) that there is _________ between X and Y.

a)$\star$ no linear relationship
b) no quadratic relationship
c) no relationship of any kind
d) a strong linear relationship
e) a strong quadratic relationship
f) a strong relationship of some kind

Problem 17. A process can fail to be stationary in various ways. For example, it can have:

1. non-constant variance
2. non-constant ACF
3. non-constant mean

where "non-constant" just means "varying with time." Differencing a non-stationary process can sometimes "cure" the problem and lead to a stationary process. Which of the above forms of non-stationarity can sometimes be cured by differencing?

a) 1 **b)** 2 **c)**$\star$ 3 **d)** 1, 2 **e)** 1, 3 **f)** 2, 3 **g)** 1, 2, 3

Problem 18. Suppose you use PROC ARIMA to fit a time series model to the series z_t and your code includes the statement

ESTIMATE P=(2,4);

What model are you fitting?

a) $z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \theta_4 a_{t-4}$
b) $z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \phi_3 z_{t-3} + \phi_4 z_{t-4} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$
c) $z_t = C + a_t - \theta_2 a_{t-2} - \theta_4 a_{t-4}$
d) $z_t = C + \phi_2 z_{t-2} + a_t - \theta_4 a_{t-4}$
e) $z_t = C + \phi_4 z_{t-4} + a_t - \theta_2 a_{t-2}$
f)$\star$ $z_t = C + \phi_2 z_{t-2} + \phi_4 z_{t-4} + a_t$

Problem 19. Suppose you wish to choose an appropriate ARIMA(p, d, q) model for the series $z_1, z_2, \ldots, z_n$. Let $y_t = (1 - B)z_t$ denote the first differences, and $w_t = (1 - B)^2 z_t$ denote the second differences. In which of the following situations would you choose the value $d = 1$?

a) y_t is stationary and w_t is non-stationary
b) z_t is stationary and w_t is non-stationary
c)$\star$ z_t is non-stationary and y_t is stationary
d) z_t stationary and y_t is non-stationary
e) y_t is non-stationary and w_t is stationary
f) z_t is stationary and w_t is stationary
g) y_t is non-stationary and w_t is non-stationary

Problem 20. Let L be the Likelihood value, k be the number of estimated parameters, and n be the number of residuals. Which of the following is the correct formula for SBC (Schwarz's Bayesian Criterion) also known as BIC (Bayesian Information Criterion)?

a) $-2\ln(L) - k\ln(n)$ **b)** $2\ln(L) - k\ln(n)$ **c)** $-\ln(L) - k\ln(n)$

d) $\ln(L) - k\ln(n)$ **e)** $-2\ln(L) - 2k$ **f)** $2\ln(L) - 2k$

g)$\star$ $-2\ln(L) + k\ln(n)$ **h)** $-\ln(L) - 2k$ **i)** $\ln(L) - 2k$

Problem 21. If you are unable to find a good ARIMA(p, d, q) model for a time series, one alternative is to _________

a) use the Box-Jenkins approach

b)$\star$ model the series as (Series) = (Trend) + (Stationary Process)

c) try differencing the series

d) try including additional AR or MA terms

e) re-express the model in backshift notation

f) check that the model satisfies the stationarity conditions

Problem 22. When you fit a regression model with sample size n and p covariates, the value of R^2 (R-squared) tends to _________ the performance of the regression, especially when _________ (Choose the pair which best completes this sentence.)

a) understate / p is large and n is small

b) understate / p is small and n is large

c)$\star$ overstate / p is large and n is small

d) overstate / p is small and n is large

e) overstate / parameter estimates and standard errors are large

f) overstate / parameter estimates and standard errors are small

g) understate / parameter estimates and standard errors are large

h) understate / parameter estimates and standard errors are small

Problem 23. For a given time series $z_1, z_2, \ldots, z_n$, the sample PACF value $\hat{\phi}_{33}$ may be approximated by _________

a)$\star$ the coefficient of z_{t-3} in a regression of z_t on $z_{t-1}, z_{t-2}, z_{t-3}$

b) the coefficient of a_{t-3} in a regression of z_t on $a_{t-1}, a_{t-2}, a_{t-3}$

c) the coefficient of a_{t-3} in a regression of a_t on $a_{t-1}, a_{t-2}, a_{t-3}$

d) the ACF value r_3 when n is large

e) the ψ-weight ψ_3 when n is large

f) the AR coefficient ϕ_3 when n is large

g) the MA coefficient θ_3 when n is large

Problem 24. For an ARMA(p, q) process, the 'dual' process is the ARMA(q, p) process obtained by interchanging the roles of the θ's and ϕ's. Which one of the following statements is true?

a) The **I**ACF of an ARMA(p, q) process is the **P**ACF of the dual process.
b) The **P**ACF of an ARMA(p, q) process is the ACF of the dual process.
c) The **P**ACF of an ARMA(p, q) process is the **I**ACF of the dual process.
d) The ACF of an ARMA(p, q) process is the **P**ACF of the dual process.
e)$\star$ The **I**ACF of an ARMA(p, q) process is the ACF of the dual process.
f) The ACF of an ARMA(p, q) process is the ACF of the dual process.

Problem 25. In regression, the Error Sum of Squares is _________

a) an unbiased estimate of the standard error
b) always smaller than the Error Mean Square
c) always greater than the Adjusted R-squared
d) used to identify cases with unusual response values
e)$\star$ the sum of the squared residuals
f) always between 0 and 1
g) expected to follow (roughly) a straight line

Problem 26. The process

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3}$$

may be written in backshift form as $\phi(B)z_t = \theta(B)a_t$. What are the polynomials $\phi(B)$ and $\theta(B)$? (Put your answers in the blanks provided below.)

$\phi(B) = \underline{1 - \phi_1 B - \phi_2 B^2}$

$\theta(B) = \underline{1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3}$

Problem 27. Suppose that

$$\tilde{z}_3 = \psi_0 a_3 + \psi_1 a_2 + \psi_2 a_1$$

where a_1, a_2, a_3 are random shocks, and ψ_0, ψ_1, ψ_2 are constants. What is $\text{Var}(\tilde{z}_3)$?

a) $\sigma_a^2(\psi_0\psi_1 + \psi_1\psi_2)$
b) $\sigma_a^2\psi_0\psi_2$
c) $\sigma_z^2(\psi_0^2 + \psi_1^2 + \psi_2^2)$
d) $\sigma_z^2(\psi_0\psi_1 + \psi_1\psi_2)$
e)$\star$ $\sigma_a^2(\psi_0^2 + \psi_1^2 + \psi_2^2)$
f) $\sigma_z^2\psi_0\psi_2$

A small data set contains 38 observations and four variables, a response variable Y and three covariates X1, X2, X3. A regression of Y on X1, X2, X3 was carried out. Some diagnostic plots from this regression are given on the following page, and the page after that gives several columns containing the data, predicted values, residuals, Cook's D (cookd), leverage, and the Studentized residuals (rstudent). Using this information, answer the following.

Problem 28. There are three fairly unusual observations in this data set. Which are they? (Write the observation numbers in the blanks provided below.)

2 13 28

Problem 29. Two of the three observations have unusual covariate values. Which are they? (Write the observation numbers in the blanks provided below.)

2 13

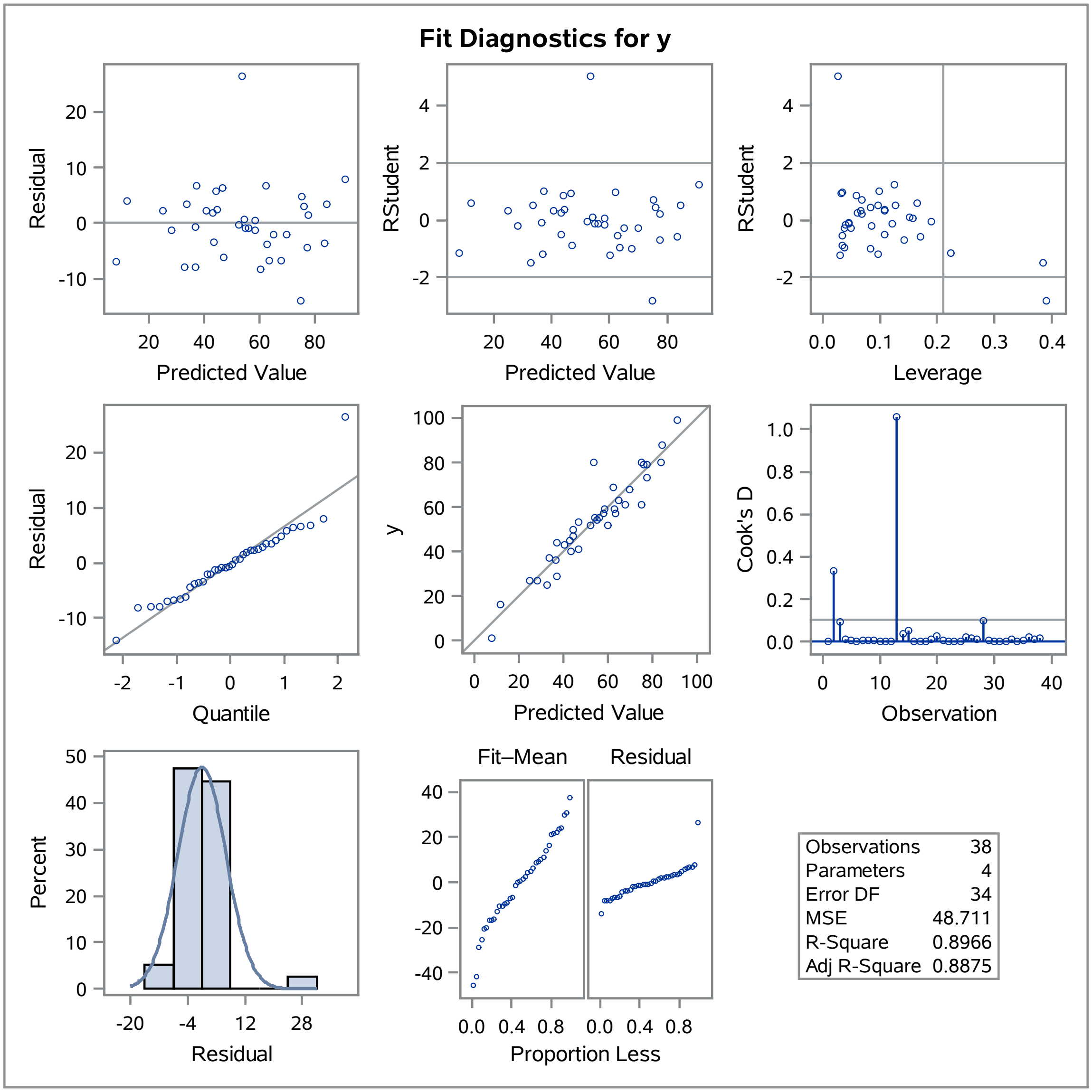
Problem 30. Two of the three observations have unusual response values. Which are they? (Write the observation numbers in the blanks provided below.)

13 28

Problem 31. One of the observations has a much greater effect on the regression model (such as on the the estimated parameters and predicted values) than the other two. Which is it? (Write the observation number in the blank provided below.)

13

The REG Procedure
Model: MODEL1
Dependent Variable: y



Obs	y	x1	x2	x3	predict	resid	cookd	leverage	rstudent
1	55	53	46	68	55.9553	-0.9553	0.00073	0.12103	-0.14388
2	25	83	97	11	32.9968	-7.9968	0.33366	0.38479	-1.48655
3	1	2	1	1	8.0003	-7.0003	0.09358	0.22404	-1.14378
4	52	50	69	59	60.2310	-8.2310	0.01141	0.03083	-1.20592
5	88	60	82	93	84.5145	3.4855	0.00745	0.09735	0.51998
6	36	35	47	30	36.7252	-0.7252	0.00014	0.04665	-0.10486
7	53	47	57	44	46.6557	6.3443	0.00741	0.03350	0.92260
8	79	77	84	85	76.0823	2.9177	0.00437	0.08399	0.43154
9	47	64	56	48	44.6134	2.3866	0.00397	0.10800	0.35738
10	27	13	37	9	24.8189	2.1811	0.00333	0.10853	0.32661
11	54	48	72	48	54.9648	-0.9648	0.00024	0.04544	-0.13943
12	45	44	45	45	43.2065	1.7935	0.00113	0.06041	0.26145
13	61	8	25	94	75.0221	-14.0221	1.05844	0.39013	-2.82427
14	29	24	55	21	36.9413	-7.9413	0.03875	0.09750	-1.20568
15	99	75	98	99	91.0549	7.9451	0.05355	0.12620	1.22683
16	68	66	81	73	69.9828	-1.9828	0.00113	0.05036	-0.28758
17	63	53	67	69	65.0204	-2.0204	0.00085	0.03767	-0.29110
18	43	51	51	40	40.7333	2.2667	0.00200	0.06605	0.33163
19	50	31	38	46	44.2192	5.7808	0.01167	0.06009	0.85087
20	44	20	52	22	37.3513	6.6487	0.02749	0.09847	1.00340
21	41	44	61	41	47.0995	-6.0995	0.00690	0.03374	-0.88625
22	57	61	74	57	58.2732	-1.2732	0.00036	0.03954	-0.18348
23	52	10	58	38	52.3408	-0.3408	0.00017	0.19022	-0.05346
24	27	35	37	23	28.2797	-1.2797	0.00085	0.08505	-0.18896
25	61	80	87	71	67.6903	-6.6903	0.02270	0.08307	-1.00110
26	16	6	12	2	12.0684	3.9316	0.01877	0.16499	0.61076
27	37	1	20	29	33.5954	3.4046	0.01002	0.12801	0.51672
28	80	46	61	52	53.5703	26.4297	0.09961	0.02634	5.02190
29	40	67	61	44	43.3533	-3.3533	0.00790	0.10873	-0.50331
30	59	17	62	48	58.5577	0.4423	0.00022	0.15784	0.06804
31	79	61	78	85	77.5999	1.4001	0.00079	0.06781	0.20483
32	59	48	68	63	62.8518	-3.8518	0.00278	0.03402	-0.55579
33	80	55	70	84	75.2273	4.7727	0.00939	0.06953	0.70364
34	55	38	80	38	54.3026	0.6974	0.00053	0.15276	0.10697
35	69	57	69	65	62.3103	6.6897	0.00855	0.03470	0.97487
36	73	51	92	72	77.4482	-4.4482	0.01985	0.14341	-0.68321
37	57	51	64	68	63.6738	-6.6738	0.00939	0.03803	-0.97421
38	80	99	99	96	83.6675	-3.6675	0.01720	0.17115	-0.57145