Economics 102_A: Analysis of Economic Data Cameron Winter 2019 U.C.-Davis
Answers to Final Exam

Version A

1.(a) 95% confidence interval directly from computer output is $(1528, 1729)$.
(b) $H_0 : \mu \geq 1700$ versus $H_a : \mu < 1700$ where claim is the alternative hypothesis.
From output $\bar{x} = 1629.2$ and $se(\bar{x}) = 51.108$.
So $t = (1629.2 - 1700)/51.108 = 70.8/51.108 = -1.139$.
Do not reject H_0 that $\mu \geq 1700$ at level .05 as $t = -1.139 > -t_{.05;1949} = -1.646$.
So do not support the claim that population mean outpatient spending is less than $1,700.
(c) `coins0` and `coins50` as these are most highly correlated with `outspend.`
(d) $b_1 = 1416.1$. Reason: $\widehat{y} = b_1 + b_2 d$ equals b_1 if $d = 0$ so $b_1 = \bar{y}$ for those with $d = 0$.
$b_2 = 1883.5 - 1416.1 = 467.4$. Reason: $\widehat{y} = b_1 + b_2 d$ equals $b_1 + b_2$ if $d = 1$ so $b_1 + b_2 = \bar{y}$ for those with $d = 1$.
(e) The graph is one that is roughly symmetric. The scale is from 0 to 10. About half the areas under the curve is left of 6.75 and about half is to the right of 6.75 (the median is 6.75).

2.(a) Outpatient spending by $8.045 \times 10 = \$80.45$.
(b) 99% conf. interval for β_2 is $b_2 \pm t_{.005;1950-2} \times s_{b_2} = 8.045 \pm 2.578 \times 4.52892 = 8.045 \pm 11.676 = (-3.63,\ 19.72)$.
(c) $H_0 : \beta_2 = 15$ against $H_a : \beta_2 \neq 15$.
$t = (b_2 - 15)/s_b = (8.045 - 15)/4.52892 = -1.536$.
$|t| = 1.536 < t_{.05;1950-2} = 1.646$.
Do not reject H_0. Conclude that at level 5% the coefficient does not differ from 15.
(d) Simplest is 0.0402 directly from correlation output.
[Alternatively $s^2_{xy} = R^2$ from regress y on x yields $r^2_{oa} = 0.0016$ and $r_{oa} = \sqrt{0.0016} = 0.04$.
Rule out -0.04 because from the regression there is positive correlation between `outspend` & age].
(e) Most likely errors are heteroskedastic. One clue is that `lnout` is symmetric so `outspend` (before taking log) is right-skewed. And we expect more variability around the line for high values of `age` than around low values.
(f) `regress outspend age, vce(robust)`

3.(a) Prediction for `outspend` is $\widehat{y}^* = b_1 + b_2 x^* = 1325.3 + 8.0453 \times 50 = 1727.6$. Outpatient spending is $1,728.
(b) $\mathrm{E}[y|x = 50] \in b_1 + b_2 x \pm t_{.025;n-2} \times s_e \times \sqrt{\frac{1}{n} + \frac{(50-\bar{x})^2}{\sum_i (x_i - \bar{x})^2}}$
$\in 1727.6 \pm 1.961 \times 2255.6 \times \sqrt{\frac{1}{1950} + \frac{(50-37.771)^2}{248054}}$
[No need for further calculation but get $\in 1727.6 \pm 1.961 \times 75.342 \in 1727.6 \pm 147.7 \in (1579.9, 1875.3)$.]
(c) Since $s^2_x = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2$ then $\sum_i (x_i - \bar{x})^2 = (n-1) \times s^2_x = 1949 \times 11.28151^2 = 248054$.
(d) The 95% confidence interval is at least $\widehat{y}^* \pm 1.96 \times s_e = \widehat{y}^* \pm 1.961 \times 2255.6 = \widehat{y}^* \pm 4423$. So the width of the confidence interval is at least 8,846, which is greater than 4,000.
(e) Four assumptions:
1. $y_i = \beta_1 + \beta_2 x_i + u_i$ where here y is Outspend and x is Age
2. u_i has mean 0 and is uncorrelated with x
3. u_i has variance σ^2 that does not vary with x
4. u_i is uncorrelated with u_j (errors for different observations are uncorrelated)

(f) Yes. [A critical value from the $T(1948)$ was used, and this requires the additional assumption of normally distributed errors (even though $T(1948)$ is very close to $N(0, 1)$)].

4.(a) Several ways to answer. A one year increase in education is associated with a 41.9 increase in outpatient spending which is about 2.5% of mean outpatient spending of 1629, so fairly small. Or a one standard deviation increase is associated with a $41.38 \times 3.066 = 126.9$ increase which is 7.7% of mean outpatient spending of 1629, so important. [This is a judgement call. In grading this what matters is the explanation].
(b) Yes, the five regressors are jointly significant at 5% since the F-statistic of overall fit equals 6.26 with $p = 0.0000 < 0.05$.
(c) Yes, the three insurance dummies (one needed to be dropped to avoid the dummy variable trap) are jointly statistically significant at 5% since `test coins25 coins50 coins95` gives $F = 7.79$ with $p = 0.0030 < 0.05$.
(d) `total=1` for all observations. So the mean of `total` is 1 and the standard deviation is 0.
(e) Yes. The only change is the choice of which of the three mutually exclusive dummies is the omitted category. So the coefficients of the other regressors, including `age`, are unchanged.
(f) Use adjusted R^2 as this penalizes larger models. Since $0.0133 > 0.0011$ the larger model fits better.
(Alternatively favor model with lower s_e^2. Since $2241.8 < 225.6$ the larger model fits better.)

5.(a) One more year of education is associated with a proportionate increase of 0.0375 or a 3.75% increase in outpatient spending.
(b) A 1% increase in age is associated with a 0.462% increase in outpatient spending.

6.(a) $\sum_{i=1}^{n} z_i = \sum_{i=1}^{4} 3 + 2i = (3 + 2 \times 1) + (3 + 2 \times 2) + (3 + 2 \times 3) + (3 + 2 \times 4) = 32$.
(b) $X_i \sim (500, 80^2)$ so $\bar{Y} \sim (500, 80^2/n) \sim (500, 6400/400) = (500, 16)$.
Expect mean 500, standard deviation $\sqrt{16} = 4$.
(c) $E[X] = 0.4 \times 1 + 0.5 \times 2 + 0.1 \times 6 = 2.0$.
$Var[X] = 0.4 \times (1-2)^2 + 0.5 \times (2-2)^2 + 0.1 \times (6-2)^2 = 0.4 + 0 + 1.6 = 2.0$.
(d) $\bar{x} = (5 + 10 + 30)/2 = 45/3 = 15$.
$\sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1) = \{(-10)^2 + (-5)^2 + (15)^2\}/2 = 350/2 = 175$.
Standard deviation is $\sqrt{175} = 13.23$.

7.(a) Missing entry (A): rmse= $\sqrt{\frac{1}{n-k} ResSS} = \sqrt{\frac{1}{18} \times 90} = \sqrt{5} = 2.236$..
(b) Missing entry (B): Use Res SS$_r$ =TSS= 360. So $F = \frac{(\text{Res SS}_r - \text{Res SS}_u)/(k-g)}{\text{Res SS}_u/(n-k)} = \frac{[360-90]/(3-1)}{90/18} = \frac{405}{15} = 27$.
Or $R^2 = 1 - \frac{90}{360} = 0.75$ so $F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{0.75/2}{0.25/18} = 27$.]
(c) Missing entry (C): $t = b/se$ so $se = b/t = 6/1.5 = 4$.
(d) The F statistic is the square of the t statistic. $F = t^2 = (4/2)^2 = 4$.

Multiple Choice Version A:

1. b **2. a** **3. c** **4. c** **5. b** **6. c** **7. d** **8. d** **9. a** **10. d**

(For **7.** a second answer given credit is **a.** because this is the answer if regress x on y).

The average GPA for the curve for this exam is 2.68.

Course grade is determined by curve based on combined course score.

Scores out of	62	
$75th$ percentile	46	(74%)
Median	37	(60%)
$25th$ percentile	32	(53%)

		C+	33 and above
A	46 and above	C	31 and above
A-	43 and above	C-	29 and above
B+	40 and above	D+	27 and above
B	38 and above	D	25 and above
B-	36 and above	D-	23 and above