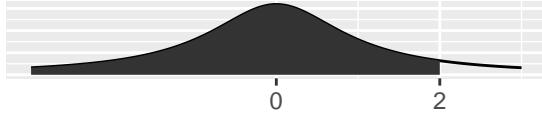


Exam 1 Practice Solutions

Econ B2000, MA Econometrics

Shay Culpepper, CCNY

Fall 2018



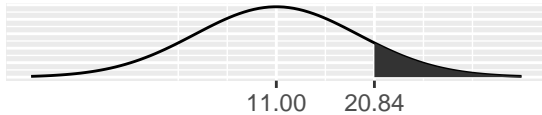
Using the Normal and Student's T to find p-values

What will be helpful in this section: `pnorm`, `qnorm`, `pt`, `qt`, and a normal distribution / student's t distribution graph to visualize. 2 ways to do it. 1) You can calculate the z score. $z = \frac{\bar{x} - \mu}{\sigma}$, or specify mean and sd in the function itself. `pnorm` and `pt` default to the lower tail.

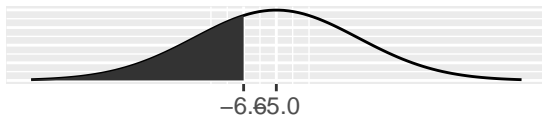
Please answer the following. You may find it useful to make a sketch

Example set 1

a. `pnorm(20.84, mean = 11, sd = 8.2, lower.tail = FALSE) ## 0.1150697`



b. `pnorm(-6.6, mean = -5, sd = 4) ## 0.3445783`



c. `pnorm(3.45, mean = 12, sd = 4.5) ## 0.02871656`

d. `pnorm(-28.82, mean = -14, sd = 7.8, lower.tail = FALSE) ## 0.9712834`

e. `pnorm(-3.7, mean = 8, sd = 9) ## 0.09680048`

f. `2 * pnorm(-14.96, mean = -10, sd = 6.8) ## 0.4657498`

g. `2 * pnorm(-6 - 12.72S, mean = -6, sd = 4.2) ## 1.361375e-05`

h. `2 * pnorm(10 - 2.8, mean = 10, sd = 6.4) ## 0.6617488`

i. `c(qnorm(0.118 / 2, mean = -6, sd = 3.5), qnorm(0.118 / 2, mean = -6, sd = 3.5, lower.tail = FALSE)) ## -11.47128 -0.5287172`

j. `c(qnorm(0.024 / 2, mean = -4, sd = 0.1), qnorm(0.024 / 2, mean = -4, sd = 0.1, lower.tail = FALSE)) ## -4.225713 -3.774287`

k. `2 * pt(4.32 / 2.7, df = 12, lower.tail=FALSE) ## 0.1355805`

l. `2 * pt(-19.11 / 9.1, df = 40) ## 0.04208202`

m. `2 * pt(-21.16 / 9, df = 29) ## 0.02572611`

Example set 7

```
a. pnorm(2.1) ## D. 0.9821
b. pnorm(-0.6) ## A. 0.2743
c. pnorm(0.3) ## C. 0.6179
d. pnorm(0.9, lower.tail = FALSE) ## A. 0.1841
e. pnorm(-0.4, lower.tail = FALSE) ## D. 0.6554
f. 2 * pnorm(-1.8) ## D. 0.0719
g. 2 * pnorm(-0.5) ## D. 0.6171
h. 2 * pnorm(-2.4) ## C. 0.0164
i. qnorm(0.324 / 2) ## C. +0.986
j. qnorm(0.390 / 2) ## A. +0.8596
k. qnorm(0.218 / 2) ## C. +1.2319
```

Example set 9

```
a. 2 * pnorm(-1.9) ## 0.05743312
b. 2 * pnorm(-1.5) ## 0.1336144
c. 2 * pnorm(-1.2) ## 0.2301393
```

Example set 10

```
a. 2 * pnorm(-3, mean = -1, sd = 1.5) ## 0.1824224
b. 2 * pnorm(-45, mean = 50, sd = 30) ## 0.00154197
c. ## 1
```

Example set 11

```
a. 2 * pnorm(1.75, lower.tail = FALSE) ## 0.08011831
b. 2 * pnorm(2, lower.tail = FALSE) ## 0.04550026
c. 2 * pnorm(1.3, lower.tail = FALSE) ## 0.193601
d. 2 * pnorm(2.1, lower.tail = FALSE) ## 0.03572884
e. c(qnorm(.1), qnorm(.9)) ## -1.281552 1.281552
f. c(qnorm(.05), qnorm(.95)) ## -1.644854 1.644854
g. c(qnorm(.025), qnorm(.975)) ## -1.959964 1.959964
```

Example set 12

```
a. pnorm(0) - pnorm(-1.75) ## 0.4599408
b. pnorm(1.75) - pnorm(0) ## 0.4599408
c. 2 * pnorm(-1.75) ## 0.08011831
d. c(qnorm(.05), qnorm(.95)) ## -1.644854 1.644854 & c(qnorm(.025), qnorm(.975)) ##
-1.959964 1.959964
```

Example set 13

```
a. pnorm(7, mean = 3, sd = 4) - pnorm(3, mean = 3, sd = 4) ## 0.3413447
b. pnorm(11, mean = 3, sd = 4) - pnorm(7, mean = 3, sd = 4) ## 0.1359051
c. 2 * pnorm(-4, mean = 3, sd = 4) ## 0.08011831
```

Statistics from given numbers (no datasets in R required)

1. *** Confidence Intervals, Hypothesis Tests

```
a.  phat <- 0.54 ## Our point estimate
    n <- 200
    critval <- -qnorm(.05)
    E <- critval * sqrt( phat * (1 - phat) / n )

    c(phat - E, phat + E)
```

```
## [1] 0.482032 0.597968
```

```
b.  phat <- 0.54
    n <- 300
    critval <- -qnorm(.05)
    E <- critval * sqrt( phat * (1 - phat) / n )

    c(phat - E, phat + E)
```

```
## [1] 0.4926694 0.5873306
```

```
c.  phat1 <- 0.54
    n1 <- 200

    phat2 <- 0.51
    n2 <- 200

    pbar <- (phat1 * n1 + phat2 * n2) / (n1 + n2)
    critval <- -qnorm(.05)
    test.stat <- (phat1 - phat2) / sqrt( pbar * (1 - pbar) * (1/n1 + 1/n2) )

    c(critval, test.stat)
```

```
## [1] 1.6448536 0.6007514
```

```
pnorm(test.stat, lower.tail = FALSE)
```

```
## [1] 0.2740028
```

```
abs(test.stat) > abs(critval)
```

```
## [1] FALSE
```

- d. Candidate X must win 2 particular states in order to win the election; the forecast says she has a 60% chance of winning each state individually. Your friend, a wannabe statistician, explains that a 0.6 chance of winning one state and a 0.6 chance of winning the other means only a $0.6 \times 0.6 = 0.36$ chance of winning both - so the “favorite” is actually not the favorite! Explain why your friend is wrong.

** I have questions on this. Does 60% mean that the polls are at 60%, or does it mean that the likelihood that the support for X is ≥ 0.51 is 60% based on their polling results?

2. *** Confidence Intervals, Minimum n (20 points)

Suppose that a particular medical treatment already improves patient outcomes by 20 (don't worry about the units for now) and it is established that the standard deviation for the population is 8. There is an improved treatment that is expected to deliver a further 10% improvement.

- a. If there were 10 patients in the trial, what would be the t-statistic, p-value, and confidence interval - assuming the new treatment works as expected? Carefully explain the null hypothesis.

```
xbar <- 22

critval <- qnorm(0.025)
t.stat <- 2 / (8/sqrt(10))

pnorm(-test.stat) ## since sigma is known

## [1] 0.2740028
```

The null hypothesis is that the improvement from this treatment improves patient outcomes by 2.

$$H_0 : \mu \leq 20$$

$$H_a : \mu > 20$$

- b. If there were 30 patients, what would be the t-stat, p-value, and confidence interval (again assuming the treatment works as expected)?

```
xbar <- 22

critval <- qnorm(0.025)
t.stat <- 2 / (8/sqrt(30))

pnorm(-test.stat) ## since sigma is known

## [1] 0.2740028
```

- c. If the company wants a p-value of 5% or lower, how many patients should they plan to have in the trial? *What is the desired margin of error??*

```
(pnorm(0.025) * 8 / 0.01) ** 2

## [1] 166446.1
```

3. Hypothesis Tests, Conditionals (20 points)

- a. Testing whether the fraction of immigrants of people making less than 15 and hour vs the fraction of immigrants of people making more than 15 an hour

```
n1 <- 14235 + 3113 + 3113 + 1824
x1 <- 3113 + 1824
phat1 <- x1 / n1

n2 <- 33150 + 662 + 5296 + 567
x2 <- (5296 + 567)
phat2 <- x2 / n2

pbar <- (x1 + x2) / (n1 + n2)
t.stat <- (phat1 - phat2) / sqrt( pbar * (1 - pbar) * (1/n1 + 1/n2))
critval <- pnorm(0.025)
p.val <- pnorm(-t.stat)
E <- critval * sqrt( phat1 * (1 - phat1) / n1 + phat2 * (1 - phat2) / n2)
point.est <- phat1 - phat2
```

$$z = 23.2265503$$

$$p\text{-value} = 1.2278041 \times 10^{-119}$$

$$E = 0.0016847$$

$$0.0720788 \leq p_1 - p_2 \leq 0.0754482$$

- a. Of immigrants, the fraction who are making \$15/hr vs who are making more than \$15/hr. In this case we'll test the proportion of immigrants making less than \$15/hr against the null hypothesis being that the proportion equals .50.

```
n <- 3113 + 1824 + 5296 + 567
phat <- (3113 + 1824) / n
p <- 0.5

critval <- -pnorm(0.005)
point.est <- phat
se.phat <- sqrt( phat * (1 - phat) / n )
se.p <- sqrt( p * (1 - p) / n )
E <- critval * se.phat
t.stat <- point.est * se.p
p.val <- pnorm(abs(t.stat)) * 2
```

$$z = 0.0021994$$

$$p\text{-value} = 1.0017548$$

$$E = -0.0024063$$

$$0.459536 \leq p \leq 0.4547233$$

- b.
- ```
n1 <- 14235 + 3113 + 1062 + 1824
x1 <- 1062 + 1824
phat1 <- x1 / n1

n2 <- 33150 + 662 + 5296 + 567
x2 <- (662 + 567)
phat2 <- x2 / n2

pbar <- (x1 + x2) / (n1 + n2)
t.stat <- (phat1 - phat2) / sqrt(pbar * (1 - pbar) * (1/n1 + 1/n2))
critval <- pnorm(0.025)
p.val <- pnorm(-t.stat)
E <- critval * sqrt(phat1 * (1 - phat1) / n1 + phat2 * (1 - phat2) / n2)
point.est <- phat1 - phat2
```

$$z = 51.1026065$$

$$p\text{-value} = 0$$

$$E = 0.0013299$$

$$0.1103247 \leq p_1 - p_2 \leq 0.1129844$$

- b.
- ```
n <- 1062 + 1824 + 662 + 567
phat <- (1062 + 1824) / n
```

```

p <- 0.5

critval <- -pnorm(0.005)
point.est <- phat
se.phat <- sqrt( phat * (1 - phat) / n )
se.p <- sqrt( p * (1 - p) / n )
E <- critval * se.phat
t.stat <- point.est * se.p
p.val <- pnorm(abs(t.stat)) * 2

```

$$z = 0.0054665$$

$$p\text{-value} = 1.0043616$$

$$E = -0.0035815$$

$$0.7049181 \leq p \leq 0.697755$$

c. $1824 / (14235 + 3113 + 1062 + 1824) \text{ ## } 0.0901453$

d. $1824 / (3113 + 1824) \text{ ## } 0.3694551$

e. $567 / (5296 + 567) \text{ ## } 0.09670817$

4. Confidence Intervals, Hypothesis Tests (20 points)

```

a. n1 <- 325
sd1 <- .1513
x1 <- -.0498

n2 <- 162
sd2 <- .1836
x2 <- .0815

critval <- pt(0.025, df = min(n1, n2) - 1)
t.stat <- (x1 - x2) / sqrt( (sd1**2)/n1 + sd2**2/n2)
p.val <- pnorm(t.stat)
E <- critval * sqrt( (sd1**2)/n1 + sd2**2/n2)
point.est <- x1 - x2

```

$$t = -7.867553$$

$$p\text{-value} = 1.8082258 \times 10^{-15}$$

$$E = 0.0085106$$

$$-0.1398106 \leq \mu_1 - \mu_2 \leq -0.1227894$$

```

b. n1 <- 112
sd1 <- .1431
x1 <- -.0349

n2 <- 75
sd2 <- .1840
x2 <- .0667

```

```
critval <- pt(0.025, df = min(n1, n2) - 1)
t.stat <- (x1 - x2) / sqrt( (sd1**2)/n1 + sd2**2/n2)
p.val <- pnorm(t.stat)
E <- critval * sqrt( (sd1**2)/n1 + sd2**2/n2)
point.est <- x1 - x2
```

$$t = -4.0342589$$

$$p\text{-value} = 2.738745 \times 10^{-5}$$

$$E = 0.0128425$$

$$-0.1144425 \leq \mu_1 - \mu_2 \leq -0.0887575$$

- c. With the R-code below, can you find other relationships? Do these differences from above seem reasonable?

```
library(quantmod)
getSymbols(c('INDPRO', 'UNRATE'), src='FRED')

## [1] "INDPRO" "UNRATE"

ip_1 <- INDPRO["1965:"]
ur_1 <- UNRATE["1965:"]
d_ip <- na.trim(ip_1 - lag(ip_1))
d_ur <- na.trim(ur_1 - lag(ur_1))
```

6. Correlations, Hypothesis Tests (20 points)

A recent research paper, looking at how much attractiveness and personal grooming affects wages, used data from The National Longitudinal Study of Adolescent Health in 2001-2.

```
a. phat1 <- 0.388
n1 <- 6074 * 0.484

phat2 <- 0.506
n2 <- 6074 * 0.506

point.est <- phat1 - phat2
pbar <- (phat1 * n1 + phat2 * n2) / (n1 + n2)
critval <- -qnorm(.05)
test.stat <- (phat1 - phat2) / sqrt( pbar * (1 - pbar) * (1/n1 + 1/n2) )
E <- critval * sqrt( phat1 * (1 - phat1) / n1 + phat2 * (1 - phat2) / n2 )

c(critval, test.stat)
## 1.6448536 0.6007514

abs(test.stat) > abs(critval)
## False
```

$$z = -4.0342589$$

$$p\text{-value} = 2.738745 \times 10^{-5}$$

$$E = 0.0128425$$

$$-0.1144425 \leq p_1 - p_2 \leq -0.0887575$$

or very well groomed; 50.6% of the females were rated that way. Is this a statistically significant difference?

- b. The study considers interrelations between physical attractiveness and grooming. People were ranked on a 4-point scale (where 1 is below average, 2 is average, 3 is above average, and 4 is very much above average) for each attribute. The full details are:

Physically

	4 Very Attractive	3 Attractive	2 Average	1 Less Attractive
4 Very well groomed	297	199	57	30
3 Well groomed	290	1169	607	54
2 Average grooming	75	788	2013	167
1 Less than average grooming	1	25	164	138

- c. Conditional on a person being ranked physically 3 or 4 in attractiveness (above average), what is the chance that they are above average (3 or 4) in grooming as well. Conditional on being above average physically, what is the chance that they are average or below average (1 or 2) in grooming? Are these statistically significantly different?

```
### Chance of above average grooming conditional on above average attractiveness
x1 <- 297 + 290 + 199 + 1169
n <- x1 + 1 + 75 + 788 + 25
phat1 <- x1 / n

### Chance of below average grooming conditional on above average attractiveness
x2 <- 1 + 75 + 788 + 25
phat1 <- x2 / n

pbar = 1
```

Personality

The study also considers the attractiveness of someone's personality (charisma), with the same 4-point scale. These data are:

	4 Very Attractive	3 Attractive	2 Average	1 Less Attractive
4 Very well groomed	326	171	60	26
3 Well groomed	416	1186	467	51
2 Average grooming	212	966	1729	136
1 Less than average grooming	11	49	184	84

- d. Conditional on having an above-average personality, what is the chance that someone has above-average grooming? Conditional on having an above-average personality, what is the chance that their grooming is at or below average? Is there a statistically significant difference?
- e. Comment on the study. If overall attractiveness is a combination of these 3 factors, is there evidence that they are gross substitutes or complements in production?

PK Robins, JF Homer, MT French (2011). "Beauty and the Labor Market: Accounting for the Additional Effects of Personality and Grooming," Labour, 25(2), pp 228-251.

7. Confidence Intervals. (30 points)

You know that a random variable has a normal distribution with standard deviation of 16. After 10 draws, the average is -12.

- What is the standard error of the average estimate? 5.0596443
- If the true mean were -11, what is the probability that we could observe a value between -10.5 and -11.5? 0.0249298

You know that a random variable has a normal distribution with standard deviation of 25. After 10 draws, the average is -10.

- What is the standard error of the average estimate? 7.9056942
- If the true mean were -10, what is the probability that we could observe a value between -10.5 and -9.5? 0.0159566

8. Confidence intervals, Hypothesis tests (15 points)

I tracked down this reference from a sign on the bus, from Tobacco Free NY. A survey of 1681 adolescents (age 11-14) in California asked if they had tried smoking and how often they went to convenience, liquor, or small grocery stores. The study finds that 452 kids rarely went to these stores and 81 had tried cigarettes; 458 kids visited these stores often (more than twice a week) and 133 had tried cigarettes. The authors assert that visiting these stores exposed the kids to more tobacco advertising.

```
n1 <- 452
phat1 <- 81/452
se1 <- sqrt(phat1 * (1-phat1) / n)
```

```
n2 <- 458
phat2 <- 133/458
se2 <- sqrt(phat2 * (1-phat2) / n)
```

```
pbar <- (133 + 81)/(452 + 458)
```

```
phat1
```

```
## [1] 0.1792035
```

```
phat2
```

```
## [1] 0.290393
```

- What is the difference in means?

```
point.est <- phat1 - phat2
point.est
```

```
## [1] -0.1111895
```

- What is the standard error of the difference in means?

```
se <- sqrt(se1 ** 2 + se2 ** 2)
```

- Is this difference statistically significant? What is the p-value? Explain.

```
t.stat <- abs(point.est / sqrt(pbar * (1 - pbar) * (1/n1 + 1/n2)))
pval <- pnorm(t.stat, lower.tail = FALSE) * 2
pval
```

```
## [1] 7.674384e-05
```

```
pval < 0.05
```

```
## [1] TRUE
```

Yes, this point estimate for the difference between these means is statistically significant. Our null hypothesis is that the difference would be 0. We have sufficient evidence to reject that hypothesis.

- d. The kids were also asked if their grades were likely to be at the level of B or below; 52 of the rare-frequency kids had belowaverage grades, while 63 of the high-frequency visitors had below-average grades. Is this difference statistically significant?

```
phat1 <- 52 / n1
phat2 <- 63 / n2

n1 <- 452
phat1 <- 81/452
se1 <- sqrt(phat1 * (1-phat1) / n)

n2 <- 458
phat2 <- 133/458
se2 <- sqrt(phat1 * (1-phat2) / n)

pbar <- (133 + 81)/(452 + 458)

phat1

## [1] 0.1792035
phat2

## [1] 0.290393
point.est <- phat1 - phat2
point.est

## [1] -0.1111895
se <- sqrt(se1 ** 2 + se2 ** 2)

t.stat <- abs(point.est / sqrt(pbar * (1 - pbar) * (1/n1 + 1/n2)))
pval <- pnorm(t.stat, lower.tail = FALSE) * 2

pval

## [1] 7.674384e-05
pval < 0.05
```

```
## [1] TRUE
```

- e. When asked about how often they had seen tobacco advertising, low-frequency visitors reported a mean of 3.1 (with standard error of 0.8) on a scale of 1-4 where 4 means “often”; high-frequency visitors reported a mean of 3.4 (with standard error of 0.8). Is this difference statistically significant?

```
xbar1 <- 3.1
se1 <- 0.8

xbar2 <- 4
se2 <- 0.8

test.stat <- (0.3) / sqrt(se1 ** 2 + se2 ** 2)
pt(test.stat, df = min(n1, n2) - 1)
```

```
## [1] 0.6044983
```

- f. Discuss the study; what else might you add? A few other things come to mind: do their parents smoke? How close is the store to their house?

Hendrick, L, N C Schleicher, E C Feighery, and S P Fortmann, (2010). "Longitudinal Study of Exposure to Retail Cigarette Advertising and Smoking Initiation," Pediatrics.

9. Confidence Intervals, Hypothesis test (15 points)

You might have missed this in the news about Alice Munro winning the Nobel, but there was a study done, showing that reading literature such as Munro and Chekov tended to make people score higher on psychological tests of Affective Theory of Mind. Consider the difference between two groups of people: either they read from a selection of literary fiction or they read non-fiction articles about non-human subjects (e.g. potatoes). They were all given a test to determine how well they could identify emotion from a picture of a person's eyes. (I'm making up some of these numbers.) The Fiction group tests at 25.6 with standard deviation of 4.38; the Non-fiction group tests at 23.5 with standard deviation of 5.17. There were 41 people in the first group and 45 in the second group.

- a. What is the difference in means?

```
point.est <- 25.6 - 23.5
```

- b. What is the standard error of the difference in means?

```
se1 <- 4.38 / sqrt(41)
se2 <- 5.17 / sqrt(44)
```

- c. Is this difference statistically significant? What is the p-value? Explain.
d. In another test, people read either literary fiction (that had won awards) or pop fiction (ie good sales but no awards). The lit fiction group scored 26.1 with standard deviation of 5.43 while the pop fiction group scored 23.7 with standard deviation of 5.08. Is this difference statistically significant?
e. Discuss the study, both in strengths and limitations.

Kidd, D C, and E Castano, (2013). "Reading Literary Fiction Improves Theory of Mind," Science.

11. Hypothesis test, standard error (15 points)

A (joking) study in the New England Journal of Medicine linked a country's per capita consumption of chocolate with the number of Nobel Prizes. It reported a regression coefficient but I got the data and did my own analysis. Five countries with the highest consumption of chocolate (UK, Belgium, Germany, Norway, Switzerland) had 19.02 Nobel Prizes per 10m people (std dev 9.0); the next five countries (USA, Finland, Denmark, Austria, Ireland) with lower chocolate consumption had 16.13 prizes (std dev 8.1).

- a. Construct a hypothesis test for whether there is no statistically significant difference between the most chocolate-consuming countries and the next group. What is the standard error of the difference? What is the test statistic? What is the p-value?

```
xbar1 <- 19.02
n1 <- 5
s1 <- 9.0
se1 <- s1 / sqrt(n1)

xbar2 <- 16.13
n2 <- 5
s2 <- 8.1
se2 <- s2 / sqrt(n2)
```

```
test.stat <- (xbar1 - xbar2) / sqrt(se1 ** 2 + se2 ** 2)
pt(test.stat, df=4)
```

```
## [1] 0.6890877
```

- b. Discuss the study. The countries that ate the most chocolate consumed 9.28 kg/person (std dev 0.54) while the next group ate merely 7.34 kg/person (std dev 1.52). Should you have eaten a chocolate bar before this exam? *F H Messerli (2012). "Chocolate Consumption, Cognitive Function, and Nobel Laureates" N Engl J Med 367: 1562-1564 October 18, 2012*

13. Hypothesis test (15 points)

A survey from eFinancialCareers found that, despite predictions from NY State that the Wall St bonus pool would drop by about 35%, a full 48% of the 911 respondents believed that their own bonuses would rise.

- a. Test the null hypothesis that the fraction of respondents expecting a bigger bonus is different from 35%. What is the p-value?

$H_0 : p \leq 0.35$ $H_a : p > 0.35$

```
phat <- 0.48
n <- 911
p <- 0.35

crit.val <- qnorm(0.975)
test.stat <- (phat - p) / sqrt(p * (1-p) / 911)
pnorm(test.stat, lower.tail = FALSE)
```

```
## [1] 9.643125e-17
```

- b. Create a 95% confidence interval for the fraction expecting a bigger bonus. What is the 90% confidence interval? The 99% interval?

```
critval <- qnorm(0.025)
E <- critval * sqrt(phat * (1 - phat) / n)
c(phat - E, phat + E)
```

```
## [1] 0.5124423 0.4475577
```

```
critval <- qnorm(0.05)
E <- critval * sqrt(phat * (1 - phat) / n)
c(phat - E, phat + E)
```

```
## [1] 0.5072264 0.4527736
```

```
critval <- qnorm(0.005)
E <- critval * sqrt(phat * (1 - phat) / n)
c(phat - E, phat + E)
```

```
## [1] 0.5226364 0.4373636
```

- c. Discuss. What other survey question might help explain this difference?

Regression Analysis from given data (no datasets in R required)

1. Fill in p-values

To investigate an hypothesis proposed by a student, I got data, for 102 of the world's major countries, on the fraction of the population who are religious as well as the income per capita and the enrollment rate of boys

and girls in primary school. The hypothesis to be investigated is whether more religious societies tend to hold back women. I ran two separate models: Model 1 uses girls enrollment rate as the dependent; Model 2 uses the ratio of girls to boys enrollment rates as the dependent. The results are below (standard errors in italics and parentheses below each coefficient):

Find the t-stat by dividing the coefficient by the standard error. Find the p-value using the first coefficient in model 1 as an example `2 * pt(-137/18, df=99)`

	Model 1	t-stat	p-value
Intercept	137 (18)	<code>round(137/18, 2)</code>	<code>2 * pt(-137/18, df=101)</code>
Religiosity	-0.585 (0.189)	<code>round(-0.585 / 0.189, 3)</code>	<code>2 * pt(-0.585 / 0.189, df=99)</code>
GDP per capita	0.00056 (0.00015)	<code>round(0.00056/0.00015, 2)</code>	<code>2 * pt(-0.00056/0.00015, df=99)</code>

	Model 2	t-stat	p-value
Intercept	1.12 (0.09)	<code>round(1.12 / 0.09, 4)</code>	<code>pt(-1.12 / 0.09, df=99)</code>
Religiosity	-0.0018 (0.0009)	<code>round(-0.0018 / 0.0009, 4)</code>	<code>pt(-0.0018 / 0.0009, df=99)</code>
GDP per capita	0.0000016 (0.0000007)	<code>round(0.0000016/ 0.0000007, 4)</code>	<code>pt(- 0.0000016/ 0.0000007, df=99)</code>

- Which coefficient estimates are statistically significant? What are the t-statistics and p-values for each?
- How would you interpret these results?
- Critique the regression model. How would you improve it?

2. Fill in p values (20 points)

Peter Gordon, in his talk at CCNY, presented results from linear regressions to explain the growth of metropolitan areas. He begins with a simple model to explain population growth from 1990-2000:

Log Population Growth 1990-2000

	Coefficient	t-stat	p-value
Constant term	-0.0229	-0.12	
Population in 1990 (log)	0.0192	1.33	<code>pt(1.33, lower.tail = FALSE, df = 67) * 2</code>
Pop. Density in 1990	-0.0504	-1.65	<code>pt(-1.65, df = 67) * 2</code>
% in manufacturing	-0.0028	-1.63	<code>pt(-1.53, df = 67) * 2</code>
R ² 0.57			

Where he also includes dummy variables for Census Regions (New England, Mid Atlantic, etc.). There are 79 observations and 67 degrees of freedom.

- a. What are the p-values for the 3 coefficients? Are they significant?

The averages and standard deviations are:

	Average	Standard deviation
Population in 1990 (log)	14.52	14.89
Pop. Density in 1990	1.80	1.02
% in manufacturing	18.69	7.75

- b. What is the predicted population growth for a metropolitan area that is exactly average?

$$pop\ growth = \beta_0 + \beta_1 (1990\ pop) + \beta_2 (1990\ pop\ density) + \beta_3 (\% \ manufacturing)$$

```

>>> r
-0.0229 + 0.0192 * 14.52 + -0.0504 * 1.80 + -0.0028 * 18.69
>>>

>>>
## [1] 0.112832
>>>

```

- c. What is the predicted population growth for a metro area that is one standard deviation above average in 1990 population?

```

-0.0229 + 0.0192 * (14.52 + 14.89) + -0.0504 * 1.80 + -0.0028 * 18.69

## [1] 0.39872

```

For a metro area one standard deviation above average in density? In manufacturing concentration? d. Give a careful explanation for why we would observe coefficients of these signs.

Statistics using Datasets (R required)

Regression Analysis using Datasets (R Required)