Exam 1 Cheatsheet

Econ B2000, MA Econometrics

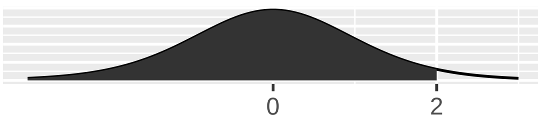
Shay Culpepper, CCNY

Fall 2018

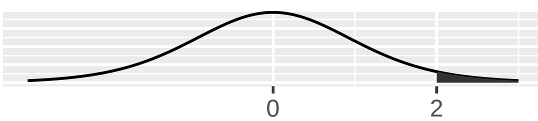
Using the Normal and Student's T to find p-values

What will be helpful in this section: `pnorm`, `qnorm`, `pt`, `qt`, and a normal distribution / students 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.

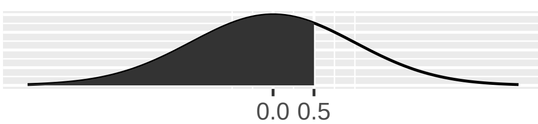
`pt(2, df = 10)` gives you the area of the shaded region in the figure below.



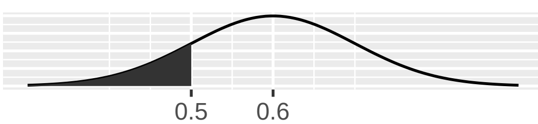
`pt(2, df = 10, lower.tail = FALSE)` gives you the area of the shaded region in the figure below.



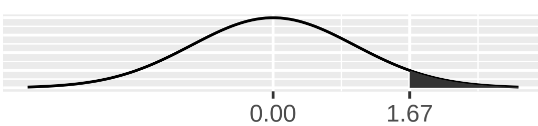
`pnorm(0.5)` gives you the area of the shaded region in the figure below.



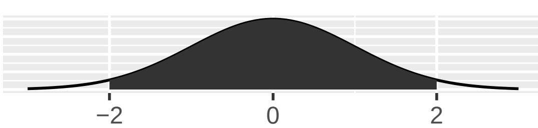
`pnorm(0.5, mean = 0.6, sd = 0.1)` gives you the area of the shaded region in the figure below.



`pnorm(1.67, lower.tail = FALSE)` gives you the area of the shaded region in the figure below.



`pnorm(2) - pnorm(-2)` gives you the area of the shaded region in the figure below.



Statistics from given numbers (no datasets in R required)

Means

These assume an unknown sigma

$$E = (t_{\alpha/2})\frac{s}{\sqrt{n}} \qquad df = n - 1 \qquad t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \qquad n = \left(\frac{z_{\alpha/2} s}{E}\right)^2 \qquad SE = \frac{s}{\sqrt{n}}$$

Difference in means

These assume variances are different and unknown

$$SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad E = t_{\alpha/2} SE \qquad df = min(n_1, n_2) - 1 \qquad t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Population proportion

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \qquad E = z_{\alpha/2} SE \qquad z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \qquad n = p(1-p)\left(\frac{z_{\alpha/2}}{E}\right)^2$$

Difference in proportions

$$SE = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \qquad E = z_{\alpha/2} SE \qquad \bar{p} = \frac{x_1 + x_2}{n_1 + n_2} \qquad z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\bar{p}\left(1 - \bar{p}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Hypothesis Test

$$\begin{array}{rrrrr}
H_0 & H_a & Test & Reject\ H_0 if... & \\
\mu \geq a & \mu < a & left - tail & test\ stat < -Z_\alpha & or\ p < \alpha \\
\mu \leq a & \mu > a & right - tail & test\ stat > Z_\alpha & or\ p < \alpha \\
\mu = a & \mu \neq a & two - tail & |test\ stat| > Z_{\alpha/2} & or\ p < \alpha
\end{array}$$

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

$$test\ stat = \frac{\beta_i}{SE_i} \qquad SE_i = \frac{\beta_i}{test\ stat} \qquad \beta_i = (test\ stat)SE \tag{1}$$

To caluclate p-value, use student's t if you have n or are given the degrees of freedom explicitly. Generally, your degrees of freedom for a beta coefficient should be $n - p$ with n being your sample size, and p being the number of parameters being predicted including the constant coefficient. If you haven't been given degrees of freedom or the sample size, you are probably safe to assume that using the normal would be sufficient.

Regularize data

```
function(x) {
  M <- max(x,na.rm = TRUE)
  m <- min(x,na.rm = TRUE)

  norm_varb <-  (x-m)/abs(M-m)
}
```

Statistics using Datasets (R required)

Regression Analysis using Datasets (R Required)