Exam 1 Solution, Math 445, Spring 2017

[1] 20 points

(a) Let μ_T = avg. proportion of 3-letter words for Twain and
μ_S = " " Snodgrass

Then our hypotheses are

$H_0: \mu_T = \mu_S$ vs. $H_a: \mu_T \neq \mu_S$

(b) To carry out a two sample permutation test in this situation we complete the following steps:

1. Calculate the difference in sample means, $\bar{X}_T - \bar{X}_S$.
2. Combine the Twain and Snodgrass data into a single vector.
3. Take a sample of size $n_T = 8$ from the vector, w/o replacement, and assign them values to Twain. Assign the remaining values to Snodgrass.
4. Calculate the difference in sample means, $\bar{X}_T^* - \bar{X}_S^*$, for this permutation. Store the value.
5. Repeat steps 3 and 4 a large # of times, say $R = 9999$.
6. Calculate the pvalue as follows:

$$\text{p-value} = \frac{\#\text{ simulated stats as/more extreme than observed} + 1}{R+1} \times 2$$

(c) The null dsn will be centered at 0, the hypothesized value

(d) $\overline{X}_T - \overline{X}_S = .231875 - .2097 = .022175$

There are 50 simulated mean differences at least as larg as .022175,
so the pvalue is given by

$$\text{pvalue} = \frac{50+1}{9999+1} = .0051$$

(e) Based on a pvalue of .0051, there is very strong evidence that the mean proportion of 3-letter words is differnt between Twain and Snodgrass.

2 9 points

(a) A test of homogeneity is appropriate here, as three samples were drawn (one in each year) and one variable is compared across the years.

(b) Let P_{2014i}, P_{2015i}, P_{2016i} be the proportions in each satisfaction group in each year.

H_0: $P_{2014i} = P_{2015i} = P_{2016i}$ $\forall$ $i = 1,2,3$ (unsatisfied, satisfied, very satisfied)

H_a: at least one of the equalities doesn't hold

(c) $E_{33} = \frac{R_3 C_3}{n} = \frac{100 \cdot 181}{300} = 60.333$

3 15 points

(a) We are considering $n = 125$ securities, so the expectations are

# Correct	0	1	2	3	4
Expected Count	7.8125	31.25	46.875	31.25	7.8125

where $E_i = np_i$ and $p_i = P(X = x_i)$

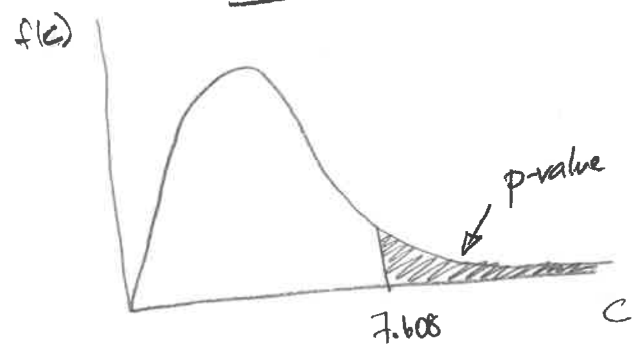
(b) $$C = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

$$= \frac{(3 - 7.8125)^2}{7.8125} + \frac{(23 - 31.25)^2}{31.25} + \frac{(51 - 46.875)^2}{46.875} + \frac{(39 - 31.25)^2}{31.25} + \frac{(9 - 7.8125)^2}{7.8125}$$

$$= 7.608$$

(c) Here, we use the χ^2_4 dsn as our reference dsn.

<u>χ^2_4</u>



(d) All of our expected cell counts are greater than 5, so Cochran's rule is satisfied, and we have a RS, so we don't have any concerns.

$\boxed{4}$ 10 points

(a) We have a theorem stating that

$$f_{max}(y) = n\left[F(y)\right]^{n-1} f(y)$$

Here $f(y) = 4y^3$, $y \in (0,1)$ and

$$F(y) = \int_0^y 4t^3 dt = t^4\Big|_0^y = y^4, \quad y \in (0,1)$$

So

$$f_{max}(y) = n\left[y^4\right]^{n-1} 4y^3 = 4ny^{4n-1}, \quad 0<y<1$$

(b) The SE of Y_{max} is the SD of the dsn of Y_{max}.

$$E(Y_{max}) = \int_0^1 y \cdot 4ny^{4n-1} dy = \int_0^1 4ny^{4n} dy = \frac{4n}{4n+1} \cdot y^{4n+1}\Big|_0^1 = \frac{4n}{4n+1}$$

$$E(Y_{max}^2) = \int_0^1 y^2 4ny^{4n-1} dy = \int_0^1 4ny^{4n+1} dy = \frac{4n}{4n+2} \cdot y^{4n+2}\Big|_0^1 = \frac{4n}{4n+2}$$

So

$$Var(Y_{max}) = E(Y_{max}^2) - \left[E(Y_{max})\right]^2$$
$$= \frac{4n}{4n+2} - \left(\frac{4n}{4n+2}\right)^2$$

and finally,

$$SE(Y_{max}) = \sqrt{\frac{4n}{4n+2} - \left(\frac{4n}{4n+2}\right)^2}$$

$\boxed{5}$ 6 points

(a) $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

(b) $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$