

LAST NAME (Print): **KEY**

Statistics 111      **Quiz 3**

1. In the 1950s, epidemiologists noticed the Jewish women had lower rates of cervical cancer, and suspected circumcision was protective. This was confirmed when Muslim women were also found to have lower rates of cervical cancer.

**Yes** Was this an observational study?

What confounding variable might explain these findings? Why?

**As we know from the homework, HPV is a cause of cervical cancer, and is spread by sexual activity. Observant Jewish and Muslim women were less likely to have sex with multiple partners, and thus less likely to contract HPV.**

2. For  $|p| < 1$ , what is  $\sum_{i=1}^{\infty} p^i$ ?  **$\frac{p}{1-p}$**
3. I draw a random sample of 8 exam scores from a large class: 28 30 20 22 23 28 29 25. Find the following:

**mode = 28                  IQR = 6                  sd = 3.66**

**The mode is the most frequent value. Both 28 and 29 are 75th percentiles, so 28.5 is the midpoint; both 22 and 23 are 25th percentiles, with midpoint 22.5; so the IQR is 28.5 - 22.5. For the sd, apply the formula on p. 5 of lecture 2 to get 3.662.**

4. **0.49** For any histogram, at least what proportion of the area must lie with 1.4 standard deviations of the mean?

**Tchebyshev; p. 6 of lecture 2.  $1 - (1/1.4)^2 = 0.4897$ .**

5. **0.84** For the standard normal distribution, what proportion of the area lies within 1.4 standard deviations of the mean?

**From the standard normal table, the area above 1.4 is 0.0808, so the area below -1.4 is also 0.0808, so  $1 - 2 * 0.0808 = 0.8384$ .**

6.  **$\int_2^3 x^{-1} dx = 0.41$**                    **$\int_1^2 \exp(x/2) dx = 2.14$**                   7.  **$\int_0^2 \int_0^y 2x dx dy = 2.67$  or  $8/3$**

From the previous quiz key, the first integral is  $\ln 3 - \ln 2 = 0.4054$ . The second is  $2 \exp(x/2)|_1^2 = 2.139$ . The third is  $y^3/3|_0^2 = 2.6667$ .

7. List only the true statements (up to 7 points): **A, G**
- A:** Gauss was greatest mathematician in history.
  - B:** The area under a standard normal curve to the right of 0.8 is 0.1977.
  - C:** A sample  $X_1, \dots, X_n$  has sd  $s$ . You transform the data so  $Y = aX + b$ . The sd of  $Y_1, \dots, Y_n$  is  $as$ .
  - D:** The  $y$ -axis in a histogram shows the number of data points with that  $x$ -value.
  - E:** In a histogram, probability is shown by the height of the bar.
  - F:** A panacea is a sugar pill given to patients in the control group to ensure blinding.
  - G:** Randomization prevents systematic differences between the treatment and control groups.