Review for Final (6/5, 3:30-6:20p, Young Hall 2200)

David Diez

Some explanations have been included; if you have a question on a solution, email me and I will provide a more thorough solution. If you think you have found an error in these solutions, please email me so I can check on it and make any necessary revisions.

(1)

(a) (number male)/(total number) = 57/117.

(b) (count up all students who are male or SS majors)/(total number) = $(25 + 30 + 27)/117 = 82/117$

(c) row proportions

(d) No. We can look at the row proportions mentioned in (c) to show that they are not independent:

$$P(SS|female) = \frac{25}{60} \neq \frac{30}{57} = P(SS|male)$$

(2) Unimodal. Skewed to the right. If we had a boxplot, we might identify some outliers at the upper end. Since the distribution is skewed, using the median and IQR would be preferable as a set of summary statistics.

(3) (a) $\leftrightarrow$ (ii); (b) $\leftrightarrow$ (i); (c) $\leftrightarrow$ (iii)

(4)

(a) 3

(b) 3. notice that 75% of the data in 3 is higher than the median of the other two distributions. additionally, each quartile of the 3rd plot is higher than in the other boxplots.

(c) 1 by comparing IQRs.

(d) 3. it has its long tail to the lower numbers, so we could say it is skewed to the left or skewed to the low end

(5) For each of these, we plot a curve, shade the appropriate region, compute the Z-score(s) (or find it in the table in (e)), then find the final solution.

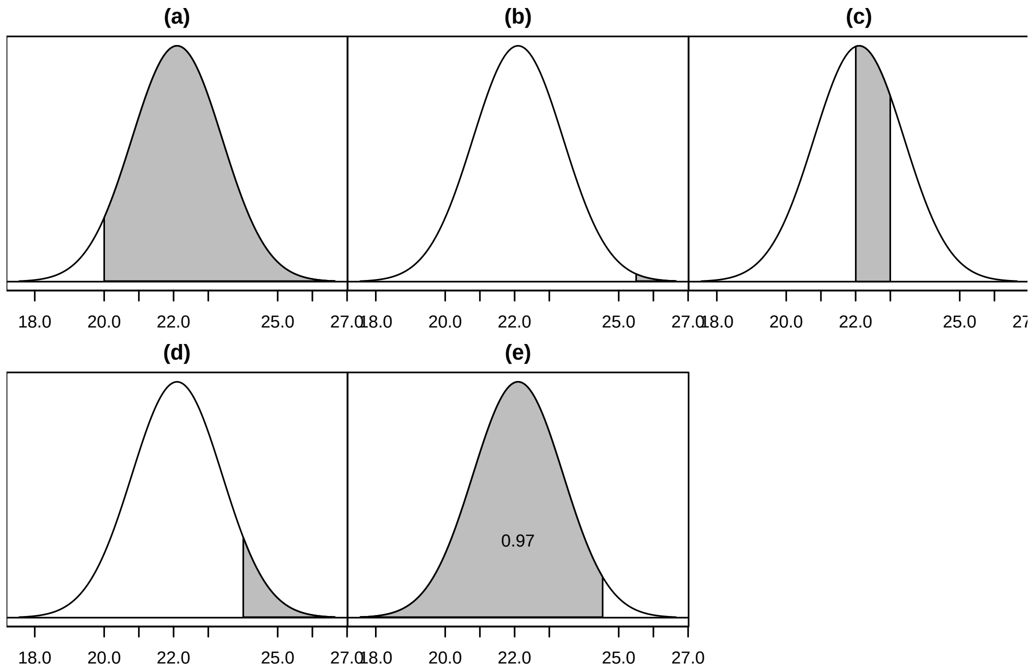


Figure 1: Plots for all 5 parts. Only the mean and, if known, the cutoff value need be labeled (they are not in the plots above). For (e), label the area known (above and/or below the cutoff).

part	look in table or compute Z?	Z	solution
(a)	compute	$Z = \frac{20-22.1}{1.3} = -1.62$	$1 - 0.053 = 0.947$
(b)	compute	$Z = \frac{25.5-22.1}{1.3} = 2.62$	0.0045
(c)	compute Z for both cutoffs	$Z_1 = \frac{22-22.1}{1.3} = -0.08$	0.469
		$Z_2 = \frac{23-22.1}{1.3} = 0.69$	$1 - 0.244$ (continued below)
(d)	compute	$Z = \frac{24-22.1}{1.3} = 1.46$	0.072 (continued below)
(e)	in neg Z table, look for .03	$Z = +1.88$ (make +)	$1.88 = \frac{x-22.1}{1.3} \to x = 24.55$ oz

(c, continued) The two solutions listed above give the area to the left of each of the cutoffs. Thus, the area between is their difference: $0.756 - 0.469 = 0.287$.
(d, continued) This problem is a binomial problem since we are looking for a precise number of occurrences and we don't care about order. The probability from the table gives $p = 0.072$. Then from the problem description, $n = 5$ and $x = 3$. Using these values and plugging into the binomial probability formula gives 0.0032.

(6)

(a) Yes since the probabilities sum to 1.

(b) MORE than 2 bags: $0.05 + 0.02 = 0.07$.

(c) One individual: $0.42 * \$0 + 0.29 * \$15 + 0.22 * \$40 + 0.05 * \$90 + 0.02 * \$140 = \$20.45$. For 20 people, we would expect about \$20.45 per person, so $\$20.45 * 20 = \$409$.

(7)

(a) $n3$ means 'not a 3'. superscripts show roll number. $P(^1n3,^3n3,^3n3,^4n3,^5n3,^63) = P(^1n3) * ... * P(^5n3) * P(^63) = \left(\frac{5}{6}\right)^5 \frac{1}{6}$. Independence was used to multiply out the probabilities.

(b) The odds of the first roll being even is 1/2, same with the second, and so on. All rolls are independent so the solution is $(1/2)^5$.

(c) Similar to (a): $\left(\frac{5}{6}\right)^3 \frac{1}{6}$

(d) We say nothing of other rolls, so this is only $P(^42 \text{ or } ^43) = P(^42) + P(^43) = 1/6 + 1/6 = 1/3$

(e) The chance of rolling a '1' is $p = 1/6$, so the expected time to a '1' is $1/p = 6$ rolls.

(8) Construct a Venn diagram (not shown here).

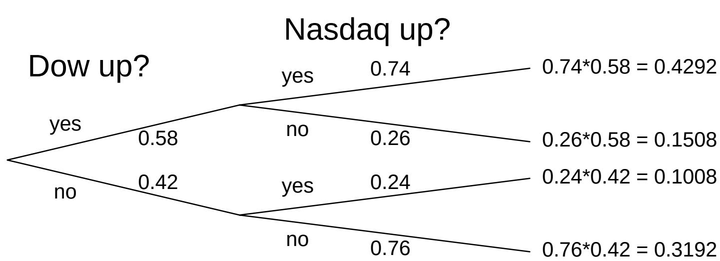
(a) 2% (like just PB) + 78% (likes both) + 11% (likes just jelly) = 91%

(b) 2% + 11% = 13%

(c) The chance of liking PB but not jelly is 0.02. This is a binomial problem (our case requires an exact number but has no specified order), so using $n = 8, x = 1, p = 0.02$, the probability is 0.139.

(9)

(a) No. The probability of the Nasdaq rising is grossly different dependent on the Dow going up (74%) or down (100% − 76% = 24%).



(b) Construct the tree diagram and add up the two cases where the Nasdaq rises:

$$0.4292 + 0.1008 = 0.53$$

(c) This is a conditional probability. What we know goes on the right, what we are trying to find the probability of goes in the first part. Then, use the definition of conditional probability.

$$\begin{aligned} P(\text{Dow up}|\text{Nasdaq up}) &= \frac{P(\text{Dow up AND Nasdaq up})}{P(\text{Nasdaq up})} \\ &= \frac{0.4292}{0.53} = 0.81 \end{aligned}$$

(10)

(a) $100\% - 27\% = 73\%$.

(b) Binomial (exact number, no order): $n = 12, x = 1, p = 0.27$. Probability: 0.102.

(c) The individuals are all independent. The order IS specified. We can break up the cases that the first through fourth are non-smokers and the last is a smoker. Multiplying out the individual probabilities yields $0.73^4 * 0.27^1 = 0.077$.

(d) Binomial: $n = 7, x = 2, p = 0.27$. Probability: 0.317.

(e) There are a range of values here, so break it up (the complement is not useful here):

$$P(0 \text{ or } 1) = P(0) + P(1) = Bin(n = 7, x = 0, p = 0.27) + Bin(n = 7, x = 1, p = 0.27) = 0.396$$

(f) Doing this directly has a lot of cases, so using the complement (and the result of (e)) gives

$$1 - P(0 \text{ or } 1) = 1 - 0.396 = 0.604$$

(11)

(a) What is the expected number of people who will show up? $\mu = n * p = 60 * 0.75 = 45$. What is the standard deviation of the number of people who will show up? $\sigma = \sqrt{np(1-p)} = 3.35$. The most important condition for normal approximations are independence and np & $n(1-p)$ are both at least 10. Independence may be a stretch but we'll let it slide for the problem. We deal with the next parts now just like normal problems with our mean and standard deviations above.

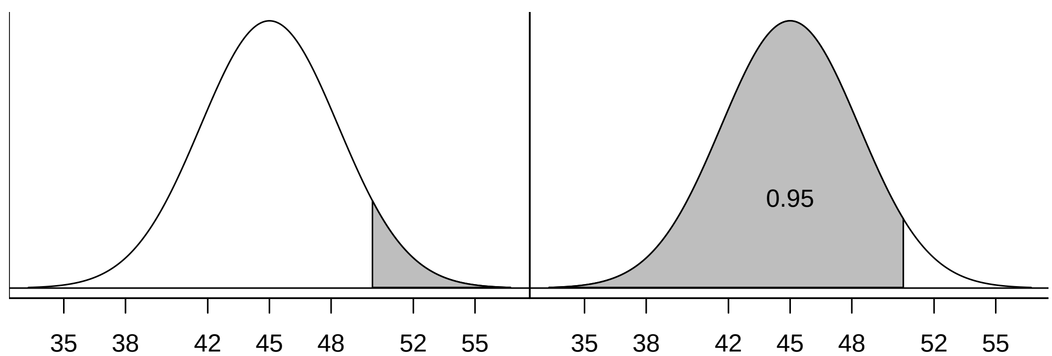


Figure 2: Shows plots for parts (b) and (c), respectively.

(b) $Z = \frac{50-45}{3.35} = 1.49 \rightarrow 0.068$.

(c) From the Z-table, $Z = 1.645$, setting that equal to our z-score: $1.645 = \frac{x-45}{3.35} \rightarrow x = 50.51$, so we need 51 seats (it only makes sense to round up).

(12)

(a) Basic normal problem. (Include a plot.) $Z = \frac{21-22.1}{1.3} = -0.85 \rightarrow 0.198$

(b) The sample mean of 5 random normal variables is still normally distributed with the same mean but with a different standard deviation (labeled SE): $SE = \frac{\sigma}{\sqrt{n}} = 0.581$. So this is just another normal problem like (a) but using a different standard deviation. $Z = \frac{21-22.1}{0.581} = -1.55 \rightarrow 0.061$. (Again, this problem should also have a plot.)

(13)

(a) $0.05 \geq ME = Z * SE = 1.96 * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. Since we don't have a $\hat{p}$, we use 0.5: $0.05 \geq 1.96\sqrt{\frac{0.5(1-0.5)}{n}} \rightarrow n = 384.16$. The sample size should be at least 385.

(b) $\hat{p} \pm Z * SE \rightarrow 0.541 \pm 1.96\sqrt{0.541(1-0.541)/385} \rightarrow (0.491, 0.591)$.

(14)

(a) F

(b) F

(c) F

(d) true

(e) F (the sample estimate is **always** in the confidence interval)

(f) true

(15)

(a) yes

(b) no

(c) $H_0 : \mu_B - \mu_A = 0$. $H_A : \mu_B - \mu_A < 0$. (Same for both.)
both plots will look like that below (not drawn with accuracy in either case but the plots just need to give some idea of what area is the p-value).

The test statistics and p-values will be different:

$$T_{paired} = \frac{-1.08 - 0}{0.21/\sqrt{6}} = -12.6 \rightarrow \text{p-value} \approx 0$$

$$T_{unpaired} = \frac{-1.08 - 0}{\sqrt{0.73^2/6 + 0.77^2/6}} = -2.49 \rightarrow \text{p-value} = 0.016$$

Both conclusions are the same:
Because the p-value is less than 0.05, we reject H_0.
Our sample provides evidence that there was an improvement in job satisfaction.

(16)

(a) Directly from the output: $(-0.015, 0.002)$.

(b) Because 0 is in the interval, we do not find a significant difference in femur length between died and survived.

(17) $H_0 : \mu_S - \mu_D = 0, H_A : \mu_S - \mu_D > 0$. $T = -1.581$. p-value = 0.058. Because the p-value is greater than 0.05, we do not reject H_0. Our sample does not provide evidence that the femurs are longer in birds that survived.

(18)

(a) Directly from the image: $\widehat{humerus} = 0.785 * femur + 0.172$.

(b) Plug it in: $\widehat{humerus} = 0.785 * 0.704 + 0.172 = 0.725$.

(c) The residual is (actual minus expected) 0.008. This observation is under-predicted.

(d) $\sqrt{0.67} = 0.82$, and it is positive since the slope is positive.

(e) r^2 represents the proportion of variation in the humerus observations that can be explained by the linear model.

(19)

(a) Find the slope after identifying the sample information (means, standard deviations, and correlation): $b_1 = r\frac{s_y}{s_x} = 0.80$. Then use point slope form, $y - \bar{y} = b_1(x - \bar{x})$, to get the equation: $\hat{y} = 0.80x + 1.76$.

(b) Plug it in: $\hat{y} = 17.0$.

(c) The residual is $13.7 - 17.0 = -3.3$, so the model over-predicted this observation. This observation may be found on the plot.

(d) Yes! Its x-value is far from $\bar{x}$ relative to other observations.

(e) Yes! Removing the point shows the line changes quite a bit. The residual of the point using the dashed line, or the LSR line not using this observation, would be very large, meaning this observation pulled the line down.

(20)

(a) gender

(b) (ii) If the student gave the same response as the previous student.

(c) only (ii) is true, since we would still not reject H_0. (a counter-example for (i) is when the p-value is 0.08.) This was definitely an observational study – gender was the explanatory variable and there is no way to control that.

(d) type II. (look to page 481 of *Intro Stats* by De Veaux, et al. for a nice chart on identifying the error type).