Test 3 Answers

1. a. T b. T

2. Choose the statistic that has the lower variability (smaller standard deviation). If the distribution is normal, this will be the sample mean. If the distribution has heavy tails, then the sample median may be a better choice.

3.. a) $n = p(1 - p)(1.96/B)^2 = .21(1 - .21)(1.96/.05)^2 = 254.93$

b) If no prior estimate of the population proportion is available, using $p = 0.50$ in the formula for *n* above would provide a sufficient sample size.

4.a. It is reasonable to use the *z* confidence interval to estimate a population mean when: (1) the data are from a random sample of the population, (2) *either* the sample size is greater than or equal to 30 *or* it is reasonable to believe that the population distribution is normal, and (3) the population standard deviation, σ, is known.

b. The *t* distribution will be an adequate approximation of the sampling distribution of the sample mean when: (1) the data are from a random sample of the population, and (2) *either* the sample size is greater than or equal to 30 *or* it is reasonable to believe that the population distribution is normal.

5.a. Two factors should be considered. First, the *statistic should be unbiased*, meaning that the statistic should have a long-term average equal to the population parameter being estimated. Second, the *standard deviation of the statistic should be as small as possible*, making typical estimates more accurate.

b. After a sample of size *n* is taken a sample proportion *p* is calculated. The sample size will be large enough to believe that the sampling distribution is approximately normal if the products $np \geq 10$ and $n(1-p) \geq 10$. Another way to say this is that the sample has produced at least 10 "successes" and at least 10 "failures".

6.a) $H_0: \pi = .35$

b) $H_a: \pi < .35$

c) A Type I error is to decide that children remember less when outside of the home environment, when in fact, they remember equally well.
A Type II error is to decide that children remember as well in the laboratory setting as in the home setting, when in fact they remember less away from home.

. 7. A p-value is the probability of observing a test statistic as or more extreme as the one observed in the sample assuming that the null hypothesis is true.

8. 1) μ = true mean PCB concentration

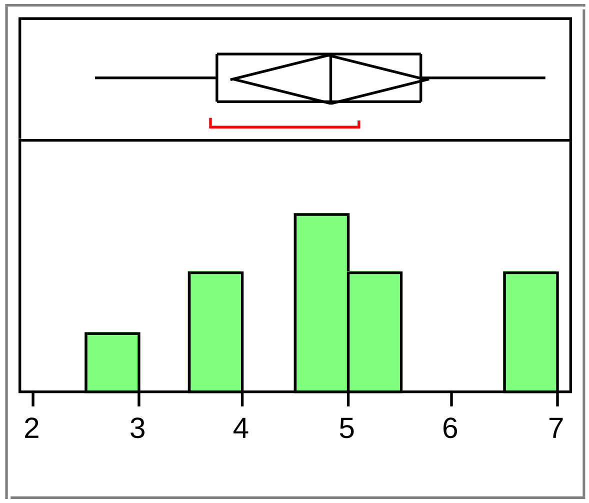
2) $H_0: \mu = 5$

3) $H_a: \mu > 5$

4) $\alpha = .05$

5) $t = \frac{\bar{x} - 5}{s/\sqrt{n}}$

6) Random sample? Given in the problem.
Normal population? Based on the sample, this seems reasonable.



7) $t = \frac{4.83 - 5}{1.301/\sqrt{10}} = -.41$ df = 9

8) P-value = $P(t > -.41) \approx .6554$

9) Since the P-value > α, we fail to reject H_0. We cannot conclude that the true mean concentration of PCB exceeds 5. That is, the sample does not give sufficient evidence that the fish should not be eaten.

9. There are 2 conditions that need to be met:

1) The data must be from a random sample from the population of interest

2) The sample size must be large (at least 30) *OR* the population must be approximately normal. To check for population normality, the sample should be graphed to see if it is plausible that it came from a normally distributed population.

10. 1) μ = true mean volume of soda dispensed (ml)

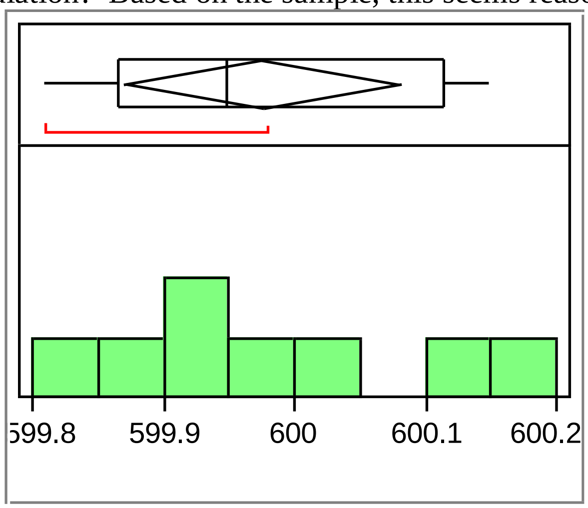
2) $H_0: \mu = 600$

3) $H_a: \mu \neq 600$

4) $\alpha = .05$

5) $t = \dfrac{\bar{x} - 600}{s/\sqrt{n}}$

6) Random sample? Given in the problem.
Normal population? Based on the sample, this seems reasonable.



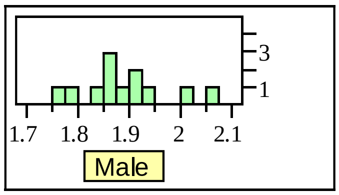
7) $t = \dfrac{599.976 - 600}{.1259/\sqrt{8}} = -.53$ df = 7

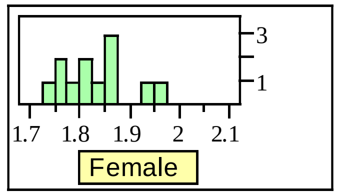
8) P-value = $2P(t < -.53) = .6102$

9) Since the P-value > α, we fail to reject H_0. We cannot conclude that the true mean volume of soda dispensed is different than 600 ml. That is, the sample does not give sufficient evidence that the machine needs adjustment.

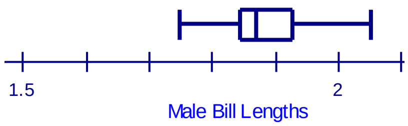
11. .   a)   It is not unreasonable to assume that the distributions of bill lengths for both male and female woodpeckers are approximately normal. The histograms and boxplots below show a reasonably symmetric distribution, possibly normal. The normal probability plots show the reasonably straight-line pattern indicating a possible normal distribution.

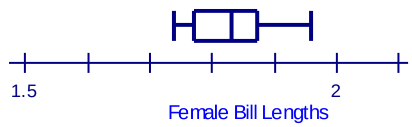
Histograms:



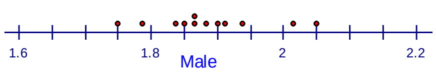


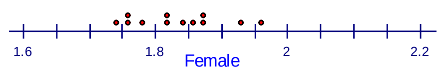
Boxplots:



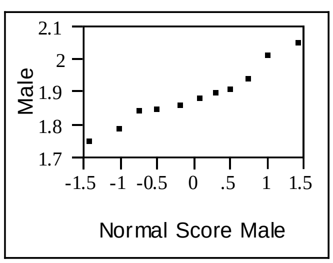


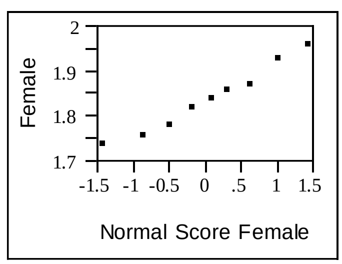
DotPlots:





Normal Probability Plots:





b) 1: Let μ_m = mean bill length for males; μ_f = mean bill length for females.

2: $H_0 : \mu_m - \mu_f = 0$

3: $H_a : \mu_m - \mu_f \neq 0$

4: Let $\alpha = 0.05$.

5: Test statistic $t = \dfrac{\bar{x}_m - \bar{x}_f}{\sqrt{\dfrac{s_m^2}{n_m} + \dfrac{s_f^2}{n_f}}}$

6: The samples were randomly selected, presumably independently. Since the sample sizes are small, part a) showed that it is reasonable to assume that the distributions of bill sizes are normal, so the distribution of sample differences should have approximately a *t*-distribution.

7: *t* = 1.67 (using formula above or a calculator).

8: *P*-value = 0.1091 (calculator) or *P*-value >0.10 (table, using 15 *df*).

9: Since the *P*-value is more than $\alpha = 0.05$, fail to reject the null hypothesis. There is not enough evidence to say that the bills of males are larger than those of females.

Let p_v = the population proportion vigorous scenery-enjoying exercisers

p_s = the population proportion of sedentary scenery-enjoying individuals

$H_0: p_v = p_s$

$H_a: p_v > p_s$

$\hat{p}_v n_v = 48;\ (1-\hat{p}_v)n_v = 38$

$\hat{p}_s n_s = 13;\ (1-\hat{p}_s)n_s = 18$

The normal c approximation is justified.

12.

$$p_c = \frac{x_v + x_s}{n_v + n_s} = \frac{48+13}{86+31} = 0.5214$$

$$z = \frac{(p_v - p_s) - 0}{\sqrt{p_c(1-p_c)\left(\frac{1}{n_v}+\frac{1}{n_s}\right)}} = \frac{(0.5581-0.4194)-0}{\sqrt{0.0924(1-0.0924)\left(\frac{1}{86}+\frac{1}{31}\right)}} = 1.326$$

$P - value = 0.092$

Since the P-value is not less than 0.05, we fail to reject the hypothesis. There is insufficient evidence, at the .05 level, that a higher proportion of vigorous excercises than sedentary individuals enjoy scenery.

13.

Let p_m = the population proportion of speech-deficient males

p_f = the population proportion of speech-deficient females

$H_0 : p_f = p_m$

$H_a : p_f < p_m$

$\hat{p}_f n_f = 22;\ (1-\hat{p}_f)n_f = 280$

$\hat{p}_m n_m = 78;\ (1-\hat{p}_f)n_m = 259$

The normal curve approximation is justified.

$$p_c = \frac{x_v + x_s}{n_v + n_s} = \frac{22+78}{302+337} = 0.1565$$

$$z = \frac{(p_v - p_s) - 0}{\sqrt{p_c(1-p_c)\left(\frac{1}{n_v} + \frac{1}{n_s}\right)}} = \frac{(0.0728 - 0.2315) - 0}{\sqrt{0.1565(1-0.1565)\left(\frac{1}{302} + \frac{1}{337}\right)}} = -5.51$$

$P - value \approx 0$

Since the P-value is less than 0.05, we reject the hypothesis. There is sufficient evidence, at the .05 level, that a higher proportion of males than females in this population have speech deficits at age 3.

b) This study is an observational study – genders were not assigned.

14. a.

Let μ_e = the population mean #correct for experimental (divided) group

μ_f = the population mean #correct for control group

$H_0 : \mu_e = \mu_c$

$H_a : \mu_e < \mu_c$

$\alpha = .05$

The approximate normality of the distributions were asserted in the problem.

$$t = \frac{(\bar{x}_e - \bar{x}_c) - 0}{\sqrt{\frac{s_e^2}{n_e} + \frac{s_c^2}{n_c}}} = \frac{(2.26 - 3.75) - 0}{\sqrt{\frac{(0.52)^2}{32} + \frac{(0.61)^2}{32}}} = -10.515$$

$P - value \approx 0$

15\. a) π_1 = the true proportion of oatmeal purchasers who choose Seabiscuit Oats
π_2 = the true proportion of oatmeal purchasers who choose Secretariat Oats
π_3 = the true proportion of oatmeal purchasers who choose Whirlaway Oats

H_0: $\pi_1 = .5,\ \pi_2 = .25,\ \pi_3 = .25$
H_a: H_0 isn't true

b) Seabiscuit: 100
Secretariat: 50
Whirlaway: 50

c) 2

d) $.045 < p < .050$

16\. a) A test of homogeneity compares the distribution of one variable in two or more populations. A test of independence tests if there is an association between two variables in one population.

b) Examples will vary. However, it is important that the sampling procedures are clearly explained. A test of homogeneity requires two or more independent random samples (measuring one variable) and a test of independence requires one random sample (measuring two variables).

17\. (a)

Expected Values for Numbers of Nests per Box by Location

Site	0 nests	1 nest	2 -3 nests
Golf	58.348	81.687	42.965
Non-golf	51.652	72.313	38.035

(b) It appears that the birds actually prefer the golf sites. There are fewer than expected nest boxes with 0 nests at the golf sites, about what was expected for the single nests, and a greater than expected count in the 2-3 nests category.

18.)

$$\text{Total} = 47 + 21 + 7 = 75$$
$$E(0 \text{ vacancies}) = 0.6065(75) = 45.4785$$
$$E(1 \text{ vacancy}) = 0.3033(75) = 22.7475$$
$$E(> 1 \text{ vacancies}) = 0.0902(75) = 6.765$$

b) $df = nCategories - 1 = (3 - 1) = 2$

c) For 2 degrees of freedom, the X^2 is not even close to significant. Thus, the hypothesis would not be rejected; the results are consistent with the theory that the proportions have not changed in the 1933-2007 time frame.