Stat 104 Assignment 7 solution

Total points: 30
Challenge Problem: 6

6.70 (2 points)
(a) " Statistically insignificant" means that the results were so small that they could have occurred simply by chance.
(b) Had the effects been large, then the statistical insignificance could be due to a small sample size. Further study with a larger sample size could procedure a result that was similar in size, but would e statistically significant.

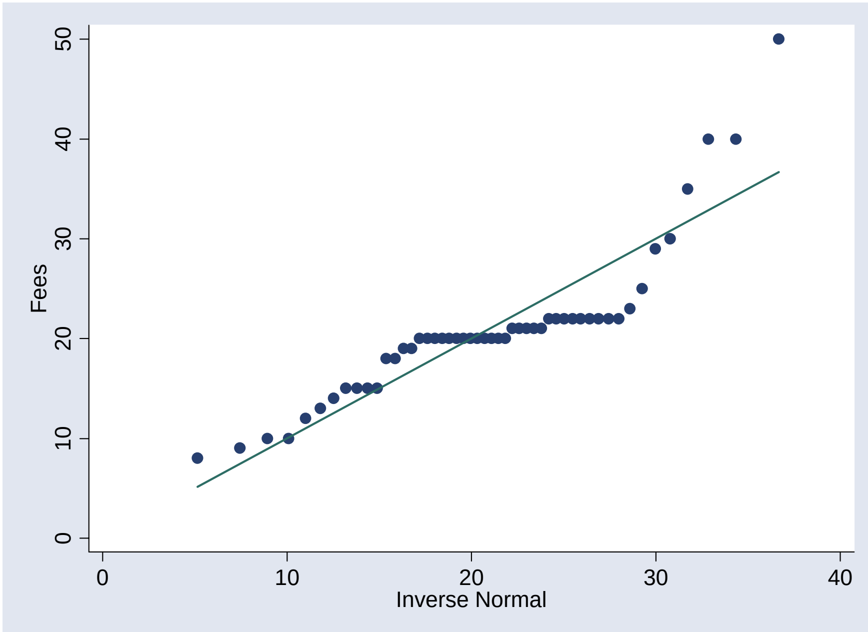
6.76(2 points)
(a) Z=1.64< 1.645, so not significant at 5%.
(b) Z=1.65 >1.645, so it is significant at 5%.

6.80(2 points)
The Bonferoni cutoff would be 0.05/12 = 0.0041666. Based on this, the 5th, the 6th and 11th are significant. (Or 0.001, 0.004, 0.002 are significant)

6.86 (3 points)
(a) The test would reject H0 when Xbar >131.46 or xbar<124.54. When mu=135, Prob(Xbar>131.46)=0.977. and Prob(Xbar<124.54)=0. So power =0.977.
(b)When mu=121, Prob(Xbar>131.46)=0 and Prob(Xbar<124.54)=0.977. Power is 0.977. So the test is reliable.
(c)The power will be higher.

6.90 (2 points)
(a) Prob(Type I error) = Prob(reject H0 when H0 is true) = Prob(X=0,1,2,3,or 5 when P0 is true) = 0.5.
(b) Prob(Type II error) = = Prob(accept H0 when H0 is false) = Prob(X=4 or 6 when P1 is true) = 0.2+0.1=0.3

7.20(3 points)
(a) Stemplot(1 point):
089
10023455555889
2000000000000111112222222235
305
400
50
Normal quantile plot(1 point):



(b) (18.73, 23.07) (1 point)

7.40(3 points)
(a) For each student, randomly select (using a coin, e.g.) which knob(right or left) should be used first.
(b)H0: $\mu=0$; Ha: $\mu>0$, where μ is the mean(left-right).
(C)Xbar=13.32, s=22.94, SE=4.59, t=2.90 with df=24. P-value is 0.0039. This is good evidence that the left-threaded knob takes longer than the right-threaded knobs

7.68(4 points)
(a) Can use any plots to compare: boxplot, q-q plot, stemplot, histogram…
(b) H0: $\mu 1=\mu 2$; Ha: $\mu 1<\mu 2$. where $\mu 1$ is the mean of the failed group and $\mu 2$ is the mean of the healthy group. The t-stat= -7.90 with 32 df. The P-value is 0. There is very strong evidence for a difference between the two groups.
(c) We CANNOT impose such a "treatment" on a firm.
(d) The confidence interval is (0.68.1.14).

7.86 (3 points)
In the example, t=0.654 with 132 df(use the table for df=100).
With the pooling, the pooled estimate of sd is 5.2679, which is between the values of s for each group. The t-stat with pooling is 0.6489. (2 points)
The P-value is not materially changed (about 50%) (1 point)

Additional problems (6 points in total)
(a) (2 points)
Ho: $\mu_E = \mu_U$; Ha: $\mu_E \neq \mu_U$.
Use a TWO-SAMPLE t test (with unequal sds or pooled sd, the results are close), t-stat =

0.8, and P-value = 0.44 . We can not reject the null hypothesis which states that there is no difference.

(b) (2 points)
d= $\mu_E - \mu_U$
Ho: d=0 ; Ha: d $\neq$ 0.
Use a ONE-SAMPLE t test for the mean difference of the pairs in the population. Now t-stat = 2.3, and P-value=0.055, it is very close to being significant. There is considerable difference, although it is not statistically significant at α=5% level.

(c) (2 point)
Numerically, it is because the SE of Xbar is higher in the one-sample test. Intuitively, in the one-sample test, we use the additional information about the relation between Europe and domestic publishing houses, and this additional information results in the big difference in the P-values.

****
Challenge Problem: (6 points, 1 point each)
1.

$$d_i = x_i - y_i \Rightarrow \frac{\sum d_i}{n} = \frac{\sum x_i - \sum y_i}{n} \Rightarrow \bar{d} = \bar{x} - \bar{y}$$

2.

$$\sigma_d^2 = \sigma^2 + \sigma^2 - 2\rho\sigma\sigma = \sigma^2(2 - 2\rho)$$

3.

$$\sigma_{\bar{d}} = \frac{\sigma_d}{\sqrt{n}} = \frac{\sigma\sqrt{2-2\rho}}{\sqrt{n}} = \sigma\sqrt{\frac{2-2\rho}{n}}$$

4.
Assume independence,

$$\sigma_{\bar{x}-\bar{y}} = \sqrt{\left(\frac{\sigma_x}{\sqrt{n}}\right)^2 + \left(\frac{\sigma_y}{\sqrt{n}}\right)^2} = \sqrt{\frac{\sigma^2 + \sigma^2}{n}} = \sigma\sqrt{\frac{2}{n}}$$

5.
Because $\sigma_{\bar{x}-\bar{y}} > \sigma_{\bar{d}}$, then the test using $\bar{d}$ and $\sigma_{\bar{d}}$ will give a bigger z-stat than the test using $\bar{x} - \bar{y}$ and $\sigma_{\bar{x}-\bar{y}}$ and assuming x and y are independent. Plug in the numbers, we will get

$$\sigma_{\bar{d}} = \sigma\sqrt{\frac{2-2\rho}{n}} = 5 * \sqrt{\frac{2-2*0.5}{20}} = 1.118$$

$$\sigma_{\bar{x}-\bar{y}} = \sigma\sqrt{\frac{2}{n}} = 5*\sqrt{\frac{2}{20}} = 1.58$$

so

$$\frac{t_d}{t_{\bar{x}-\bar{y}}} = \frac{\bar{d}/\sigma_{\bar{d}}}{(\bar{x}-\bar{y})/\sigma_{\bar{x}-\bar{y}}} = \frac{\sigma_{\bar{x}-\bar{y}}}{\sigma_{\bar{d}}} = 1.58/1.18 = 1.4$$

The bigger z-stat will be 40% larger than the smaller z-stat.

6 No. when $\rho < 0$, $\sigma_{\bar{x}-\bar{y}} < \sigma_{\bar{d}}$, the test using $\bar{d}$ and $\sigma_{\bar{d}}$ will give a smaller z-stat.