Stat13 Homework 8 -Suggested Solutions

Question 8.2:

In this observational study, the effect of implants on illness is confounded with the effects on illness of smoking, drinking heavily, using hair dye, and having an abortion.

Question 8.3:

Part (a) :

The explanatory variable is whether or not a woman has had breast implants.

Part (b):

The response variable is illness (whether or not one is ill).

Part (c):

The observational units are individual women.

Question 8.8:

It appears that the digestive problems were caused by the placebo effect. People feared that fluoridation of their drinking water would cause health problems and this fear lead to digestive problems when, in fact, fluoride was not yet being added to the water.

Question 8.19:

This is a bad proposal because treatment differences would be confounded with differences between the litters. If two treatments appeared to be different, we would not know whether this was due to a true treatment effect or due to differences in the litters.

Question 8.28:

This is a non-sampling error, because it would occur even with a census of the entire database.

Question 9.2:

Part (a) :

The standard deviation of the nine sample differences is given as 59.3. The standard error is:

$$SE_{\bar{d}} = \frac{s_d}{\sqrt{n_d}} = \frac{59.3}{\sqrt{9}} = 19.77$$

Part (b):

H_0: The mean weight gains on the two diets are the same $(\mu_1 = \mu_2)$.

H_A: The mean weight gains on the two diets are different $(\mu_1 \neq \mu_2)$.

$$t_S = \frac{22.9}{19.77} = 1.158$$

With df = 8, Table 4 gives $t_{0.20} = 0.889$ and $t_{0.10} = 1.397$. Thus, $0.20 < P < 0.40$ and we do not reject H_0. There is insufficient evidence ($0.20<P<0.40$) to conclude that the man weight gains on the two diets are different.

Part (c):

$$CI = 22.9 \pm (1.860)(19.77) = (-13.90, 59.70)$$

Part (d):

We are 90% confident that the average steer gains somewhere between 59.7 pounds more and 13.9 pounds less when on Diet 1 than when on Diet 2 (in a 140-day period).

Question 9.22:

Let p denote the probability that the Northern member of a pair will dominate in more episodes than the Carolina.

H_0: Dominance is balanced between the subspecies $(p = 0.5)$.

H_A: One of the subspecies tends to dominate the other $(p \neq 0.5)$.

$N_+ = 8$, $N_- = 0$, $B_S = 8$. Looking under $n_d = 8$ in Table 7, we see that the rightmost column with a critical value less than or equal to 8 is the column headed 0.01 (for a nondirectional alternative), and the next column is headed 0.002. Therefore, $0.002<p<0.01$. There is sufficient evidence ($0.002<p<0.01$) to conclude the Carolina subspecies tends to dominate the Northern.

Question 9.23:

$$p = 2(0.5^8) = 0.0078125$$

Question 9.33:

H_0: Alcoholism has no effect on brain density.

H_A: Alcoholism reduces brain density.

The differences tend to be negative, which is consistent with H_A.

The absolute values of the differences are 1.2, 1.7, 0.5, 4.7, 3.3, 0.4, 2.7, 1.8, 0.1, 0.3 and 1.4.

The ranks of the absolute differences are 5, 7, 4, 11, 10, 3, 9, 8, 1, 2 and 6.

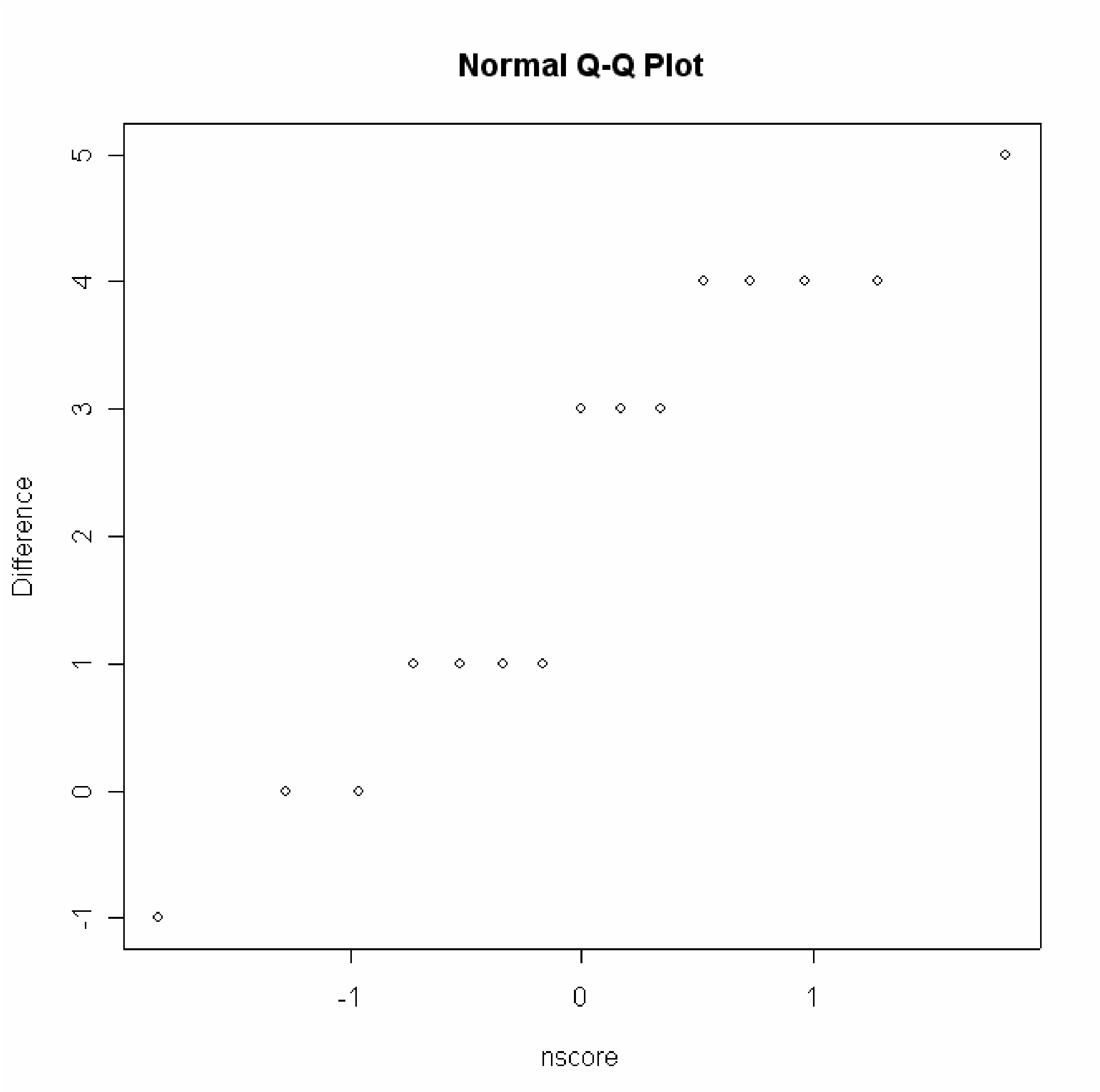
The signed ranks are -5, -7, -4, -11, -10, 3, -9, -8, -1, 2 and -6.

Thus, $W_+ = 3 + 2 = 5$ and $W_- = 5 + 7 + 4 + 11 + 10 + 9 + 8 + 1 + 6 = 61$.

$W_S = 61$ and $n_d = 11$; reading Table 8 we find $0.001 < \text{p-value} < 0.005$ and H_0 is rejected. There is strong evidence ($0.001<\text{p-value}<0.005$) to conclude that alcoholism is associated with reduced brain density. This was an observational study, so drawing a cause-effect inference is risky. We should stop short of saying that alcoholism reduces brain density.

Question 9.44:

It must be reasonable to regard the differences as a random sample from a normal population. We must trust the researchers that their sampling method was random. The normality condition can be verified with a normal probability plot. The plot below is fair linear (although the plateaus show that there are several differences that have the same value) which supports the normality condition.



Chapter 10 – Not Graded

NOTE: Chi-Square Table in SOCR is different from the textbook. SOCR shows the area to left and textbook shows tail probabilities

10.4: Non-directional
H_0: Timing of births is random (Pr(weekend) = 2/7)
H_A: Timing of births is not random (Pr(weekend) not= 2/7).

	Weekend	Weekday
Observed	216	716
Expected	266.29	665.71
Difference	-50.29	+50.29

Chi-Square = Sum of $(O-E)^2/E = (-50.29)^2/266.29 + (+50.29)^2/665.71 = 13.3$

With df = 1, http://socr.stat.ucla.edu/Applets.dir/OnlineResources.html#Tables, we get the area=.995. To get the non-directional tail, we take 1-.995 and get P-value < .005. There is sufficient evidence to conclude that the timing of births is not random.

10.5 Non-directional. Let WF and DF denote white and dark feathers; let SC and LC denote small and large comb.

H_0: The model is correct; that is, Pr(WF,SC) = 9/16, Pr(WF,LC) = 3/16, Pr(DF,SC)=3/16, Pr(DF,LC)=1/16.
H_A: The model is incorrect; that is, Probabilities are not as specified by H_0.

	OBS	EXP
WF,SC	111	106.875
WF,LC	37	35.625
DF,SC	34	35.625
DF,LC	8	11.875

Chi-square = Sum of $(O-E)^2/E = 1.55$, with df = 3, (http://socr.stat.ucla.edu/Applets.dir/OnlineResources.html#Tables. We get area between 0.25 and 0.50. To get the non-directional tail, we take 1-0.25=0.75 and also 1-0.50=0.50. The result is 0.5< P-value <.75. We do not reject H_0. There is little or no evidence (0.5< P-value <.75) to conclude that the model is incorrect; the evidence is consistent with the Mendelian model.

10.6a: Non-directional, n = 1000

	OBS	EXP	DIFF
BOY	510	500	10
GIRL	490	500	-10

Chi-square = 0.2+0.2 =0.4. With df = 1, (http://socr.stat.ucla.edu/Applets.dir/OnlineResources.html#Tables) shows .50<P-value < .75

10.6b: Non-directional, n = 5000
Shortcut: Chi-square=5*0.4=2. With df = 1, (http://socr.stat.ucla.edu/Applets.dir/OnlineResources.html#Tables). We get area between 0.75 and 0.90. We take 1-0.75=0.25 and also 1-0.90=0.10. The results is.1< P-value < .25

	OBS	EXP	DIFF
BOY	2550	2500	50
GIRL	2450	2500	-50

10.6c: Non-directional, n = 10000
Shortcut: Chi-square=10*0.4=4 With df = 1, Table 9 (http://socr.stat.ucla.edu/Applets.dir/OnlineResources.html#Tables) shows .025< P-value < .05

10.11: Directional.
H_0: The men are guessing (Pr(correct) = 1/3)
H_a: The men have some ability to detect their partners (Pr(correct) > 1/3)

	Observed	Expected
Correct	18	12
Wrong	18	24
Total	36	36

Directionality check: OK to proceed with finding p-value.
Chi-Square statistic = 4.5. With df = 1, http://socr.stat.ucla.edu/Applets.dir/OnlineResources.html#Tables gives non-directional .025 < P-value < .05. We divide in half for a directional test and get 0.0125 < P-value < .025 and we reject H_0 if alpha was=.05. Note that no alpha level was specified, but a P-value less than 0.025 is generally considered to be small.

10.17: Directional.
table is

	striped	all red	
alive	65 (70.31)	23 (17.69)	TOTAL = 88
dead	98 (92.69)	18 (23.31)	TOTAL = 116
	TOTAL = 163	TOTAL = 41	TOTAL = 204

Directionality check: OK to proceed with finding p-value.

Null is that there is no difference in the survival rates for the two types, and alternative is that the mimic form (all red) survives more than the striped kind. Test stat is chi-sq = [(65 - 70.31)^2/70.31] + [(98 - 92.69)^2/92.69] + [(23 - 17.69)^2/17.69] + [(18 - 23.31)^2/23.31] = 0.40 + 0.30 + 1.59 + 1.21 = 3.50
From SOCR, we get areas of .90 and .95 which corresponds to non-directional tails of .10 and .05. Again, since alternative is one-tailed, we half to get p-values: 0.025 < P-value < 0.05.

Since P-value $\leq \alpha$, we conclude that the mimic form of *P. cinereus* seem to survive more successfully that the red-striped. (df = 1)

10.22a: Directional.
H_0: *E. coli* had no effect on tumor incidences. p1 = p2
H_a: E. coli increased tumor incidences. p2 > p1
α = .05, Df = 1

	Germ-free	E. coli	
Tumors	19 (21.34)	8 (5.66)	27
No tumors	30 (27.66)	5 (7.34)	35
Total	49	13	62

Directionality check: OK to proceed with finding p-value.
chi-sq = 2.17. Since, chi-sq_.20 = 1.64 and chi-sq_.10 = 2.71

Multiply by half because H_a is directional: therefore, $.05 < P < .125$
We do not reject H_0.There is insufficient evidence ($.05 < P < .125$) to conclude that *E. coli* increases the number of tumors in mice.

10.22b: Directional.
If the percentages stay the same but the sample sizes double, then the O (Observed) and E (Expected) values double. Also (O-E) doubles, which means that $(O-E)^2$ is four times larger. But when divided by a doubled E, we get that $(O-E)^2 / E$ is doubled. So the Chi-square statistic is doubled and =4.34. Then H_0 is rejected because $.0125 <$ P-value $< .025$.

Similarly, if the samples were to triple, then the Chi-square statistic would triple and = 6.51. Then $.005 <$ P-value $< .0125$ and, of course, H_0 is rejected.

This makes sense. Here is an example. Given the null hypothesis that the coin is fair is really true, the true Pr (heads)=0.50. The probability of getting an extreme number of heads [for example, toss a coin only 4 times and get 3 (75%) heads] is greater than if you had tossed the coin 100 times. Let's say, you tossed a coin 100 times and got 75 (75%) heads. Then the probability of getting this extreme number of heads is highly unlikely if the coin was really fair. Remember, the p-value (probability) is given that null hypothesis is actually true.

10.35: Non-directional.
p1 = Pr{HP / MP} and p2 = Pr{HP / MA}. Null is that p1 = p2, and alternative is that p1 and p2 differ. Chi-square = 7.96.
http://socr.stat.ucla.edu/Applets.dir/OnlineResources.html#Tables
P-value < 0.005, and, since P-value < alpha, reject the null and conclude that there is an association (dependence) between the species. Data suggests repulsion. 47.3 % = p1-hat < p2-hat = 70.8%. (df = 1)

10.37a:
Pr {Yes|A} : 111/513 = 0.21637 = 21.637%
Pr {Yes|B} : 74/515 = 0.1437 = 14.37%

10.37b: Pr {A|Yes} : 111/185 = 0.60 = 60%
Pr {A|No} : 402/843 = 0.4767 = 47.67%

10.73: Directional.
Let **p** denote the probability that the uninfected mouse in a cage becomes dominant.

H0: Infection has no effect on development of dominant behavior (p = 1/3)
HA: Infection tends to inhibit development of dominant behavior (p > 1/3)

Uninfected mouse

Dominant	NotDominant
15(10)	15(20)

Directionality check looks okay. Chi-square statistic = 3.75. With df = 1, we get non-directional 0.05< P-value < 0.10. For non-directional, 0.025 < P-value < 0.05. We reject H_o. There is sufficient evidence (0.025 < P-value < 0.05) to conclude that infection tends to inhibit development of dominant behavior.

10.87: The hypotheses are

H0: Type of treatment does not affect survival
HA: Type of treatment affects survival

table is		Zidovudine	Didanosine	Both	Total
	Died	17 (11.29)	7 (11.50)	10(11.21)	34
	Survived	259(264.71)	274(269.50)	264(262.79)	797
	Total:	276	281	274	831

Chi-square statistic is 4.98; df = 2
http://socr.stat.ucla.edu/Applets.dir/OnlineResources.html#Tables
we have .025< P-value < .05 and we reject H0. At the .10 level, there is sufficient evidence (.025 < P-value < .05) to conclude that type of treatment affects survival.

10.96:
Chi-Square Test

Expected counts are printed below observed counts

	N	M	H	Total
A	18	11	4	33
	12.52	11.38	9.10	
P	4	9	12	25
	9.48	8.62	6.90	
Total	22	20	16	58

Chi-Sq = 2.402 + 0.013 + 2.861 +
3.170 + 0.017 + 3.777 = 12.238
DF = 2, P-Value <.005
Ho= no relationship between smoking and atrophied *villi*
Ha=There is a relationship between smoking and atrophied *villi*

Given that the P-Value is less than the significant value of .05, there is sufficient evidence (P-Value <.005) to conclude that, H_a = There is a relationship between smoking and atrophied *villi*. Therefore H_o is rejected.

10.96(b)

	N	M	H
A	18	11	4
P	4	9	12
Total	22	20	16
% of V	18%	45%	75%

10.96(c)
Chi-square does not show that the percentage with atrophied villi increases as smoking level increases.