Use t-table because σ is unknown we only have s.

Practice Problems

1. A random sample of 172 marketing students was asked to rate on a scale from 1 (not important) to 5 (extremely important) health benefits as a job characteristic. The sample yielded a mean rating of 3.31 and a standard deviation of 0.70. Test at the 1% level of significance the claim that the population mean rating is bigger than 3.0.

$\bar{X} = 3.31$, $s = 0.70$, $n = 172$, $\alpha = 0.01$, $\mu_0 = 3.0$

- $H_0: \mu = 3.0$ vs. $H_a: \mu > 3.0$

- $$t_{obs} = \frac{3.31 - 3.00}{0.70/\sqrt{172}} = 5.80$$

- p-value:

- p-value $< \alpha = 0.01$

5.80 is off the chart, $\Rightarrow$ p-value < 0.0005

$\Rightarrow$ reject H_0 $\Rightarrow$ Statistically Significant evidence at the 0.01 level to conclude that the pop. mean rating is bigger than 3.0

proportions

2. It is important for airlines to follow the published scheduled departure times of flights. Suppose that one airline that recently sampled the records of 246 flights originating in Orlando found that 10 flights were delayed for severe weather, 4 flights were delayed for maintenance concerns, and all the other flights were on time. Estimate the percentage of on-time departures using a 98% confidence interval. How would the interval change if you used Wilson's estimate?

Check $n\hat{p} = 232 > 10$

$n(1-\hat{p}) = 14 > 10$

$n = 246$, $x = 246 - 14 = 232$, $\hat{p} = \frac{x}{n} = \frac{232}{246} = 0.9431$

$C = 98 \Rightarrow 1 - \alpha = .98 \Rightarrow \alpha = 0.02 \Rightarrow \alpha/2 = 0.01$

$\Rightarrow z^* = 2.33$

a 98% CI:

$$\left[\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] = \left[0.9431 \pm 2.33 \cdot \sqrt{\frac{0.9431\,(1-.9431)}{246}}\right]$$

$\Rightarrow [.9086, 0.9775] \Rightarrow$ length ≈ 0.0689

Wilson's estimate: $\tilde{p} = \frac{x+2}{n+4} = \frac{234}{250} = 0.936$

CI: $\left[0.936 \pm 2.33 \cdot \sqrt{\frac{0.936 \cdot (1-.936)}{250}}\right]$

$= [0.8999, 0.97206] \Rightarrow$ length ≈ 0.07216

Using Wilson's estimate yields a slightly wider interval

3. A fast-food chain decided to carry out an experiment to assess the influence of advertising expenditure on sales. Different relative changes in advertising expenditure, compared to the previous year, were made in 8 regions of the country, and resulting changes in sales levels were observed. The data yield the following least squares regression line

$$\text{increase in sales} = 3.2958333 + 0.5391026\,\text{increase in advertising expenditure}$$

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	3.2958333	1.207511	2.73	0.0342
Increase in advertising expenditure	0.5391026	0.153664	3.51	0.0127

Slope b

estimate b

SE_b

t_{obs}

(a) Using the output given above, find a 90% confidence interval for the slope of the population regression line. $C = 90\% \Rightarrow (1-\alpha) = .90 \Rightarrow \alpha = .10$

$\Rightarrow \alpha/2 = 0.05 \Rightarrow t^* = 1.943$, $df = n-2 = \underline{\underline{6}}$

a 90% CI:

$b \pm 1.943 \cdot SE_b \Rightarrow [0.5391 \pm 1.943 \cdot 0.153664]$

$\Rightarrow [0.2405,\ 0.8376]$

(b) Based on an appropriate hypothesis test, can you conclude that increased advertising expenditures are **positively** related to increased sales? Carefully carry out all necessary steps for a hypothesis test and state your conclusions. $\alpha = 0.05$

- $H_0: \beta = 0$ vs. $H_a: \beta > 0$

- $t_{obs} = \dfrac{b}{SE_b} = \dfrac{0.5391}{0.15364} = 3.51$ (See above output)

- p-value: $0.0127 \div 2$

  $= 0.00635$ (because this is only a <u>one-sided</u> alternative $H_a: \beta > 0$

$\Rightarrow$ p-value $< \alpha = 0.05 \Rightarrow$ able to reject at the 0.05 level

$\Rightarrow$ There is statistically significant evidence at the 0.05 level to conclude that increased advertising expenditures are positively related to increased sales?

Use z^* because σ is known

4. A college admissions officer for an M.B.A. program has determined that historically applicants have undergraduate grade point averages that are normally distributed with standard deviation of 0.45. From a random sample of 25 applications from the current year, the sample mean grade point average is 2.90. $n = 25$, $\sigma = 0.45$, $\bar{x} = 2.90$

   (a) Find a 94% confidence interval for the population mean.

$C = 94 \Rightarrow (1-\alpha) = .94 \Rightarrow \alpha = 0.06 \Rightarrow \alpha/2 = 0.03 \Rightarrow z^* = 1.88$

$\Rightarrow \left[\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}\right] \Rightarrow \left[2.90 \pm 1.88 \cdot \frac{0.45}{\sqrt{25}}\right]$

$\Rightarrow [2.7308, 3.0692]$

   (b) Based on the sample results, a statistician computes for the population mean a confidence interval extending from 2.81 to 2.99. Find the confidence level associated with this interval.

length of CI $= 2.99 - 2.81 = 0.18$

$\Rightarrow$ margin of error $m = z^* \frac{\sigma}{\sqrt{n}} = 0.09$

$\Rightarrow 0.09 = z^* \cdot \frac{0.45}{5} \Rightarrow z^* = 1 \Rightarrow \alpha/2 = 0.1587$

$\Rightarrow \alpha = 0.3174 \Rightarrow \approx 68\%$ CI

5. Suppose Brent Matthews, manager of a Sam's Club in Tennessee wants to estimate the mean number of gallons of milk that are sold during a typical weekday. Brent checked the sales records for random sample of 16 days and found the mean number of gallons sold is 150 gallons per day, the sample standard deviation is 12 gallons. With 95% confidence estimate the number of gallons that Brent should stock daily.

$n = 16$, $\bar{x} = 150$, $s = 12$, $C = 95 \Rightarrow (1-\alpha) = .95 \Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025 \Rightarrow t^* = 2.131$, $df = n-1 = 15$

$\Rightarrow \left[\bar{x} \pm t^* \cdot \frac{s}{\sqrt{n}}\right]$

$= \left[150 \pm 2.131 \cdot \frac{12}{\sqrt{16}}\right] \Rightarrow [143.607, 156.393]$

Needed to assume that sampling distribution of $\bar{X}$ was approximately normal.

6. The student government association at a university wants to estimate the percentage of the student body that supports a change being considered in the academic calendar of the university for the next academic year. How many students should be surveyed if a 90% confidence interval is desired and the margin of error is only to be 2%?

as we don't know anything about p or $\hat{p}$, ~~the~~ we have to assume the most conservative p which is 0.5

$C = 90 \Rightarrow (1-\alpha) = .9$

$\Rightarrow \alpha = .1 \Rightarrow \alpha/2 = .05$

$\Rightarrow z^* = 1.645$

$m = 0.02$

$$n \geq \left(\frac{z^*}{m}\right)^2 \cdot p(1-p)$$

$$= \left(\frac{1.645}{0.02}\right)^2 \cdot 0.5(1-0.5)$$

$$= 1691.26$$

$\Downarrow$ 1692 Students

7. **Supermarket shelf space.** The general manager of a supermarket chain is interested in predicting the sales of a product based on how much shelf space is allotted to the product. He records the number of boxes of detergent sold during 1 week in 25 stores in the chain along with the shelf space (in inches) allotted to the detergent.

   (a) The store manager asks the summer intern who is helping him with the statistical analysis to fit a simple linear regression model of sales on shelf space. The results (from JMP) are shown in Figure 1. Write down the theoretical model that corresponds to the analysis performed in JMP.

   $Y = \beta_0 + \beta_1 X + \varepsilon$ where Y = sales; X = shelf space

   Alternatively, we have $\mu_Y = \beta_0 + \beta_1 X$.

   (b) Using the information provided by JMP, write down the fitted model.

   $\hat{Y} = 239.75 + 1.14X$

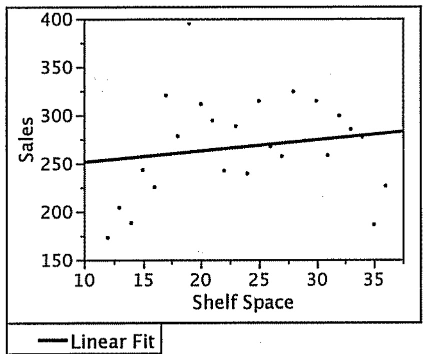
   (c) Looking over these results, the store manager concludes that there is no effect of the shelf space on the price of the detergent. How did he come to this conclusion?

   He tested the hypotheses: $H_0: \beta_1 = 0$ vs. $H_a: \beta_1 \neq 0$ and used $t = 0.80$, with 24 d.f. The p-value $= 0.4318$ which is larger than any reasonable significance level, hence fail to reject H_0. This indicates that there is no evid. of lin. rel. between sales & shelf space.

   (d) List any arguments you have against the analysis that was performed in part (a). *Hint: check whether the assumptions are satisfied or not.*

   From the residual plot (first plot in Figure 2) we see there is a quadratic pattern, indicating that the linear assumption was not satisfied. Hence the SLR model is not appropriate here, and the conclusions are not valid.

Figure 1: JMP output for the model in the **Shelf space** problem



Linear Fit

Summary of Fit

RSquare	0.02709
RSquare Adj	-0.01521
Root Mean Square Error	51.53602
Mean of Response	267.2
Observations (or Sum Wgts)	25

Analysis of Variance

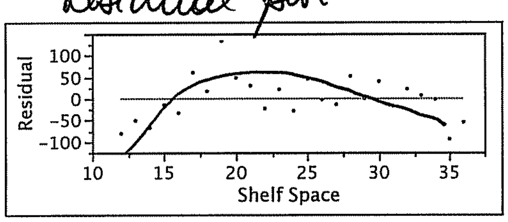
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1700.899	1700.90	0.6404
Error	23	61087.101	2655.96	Prob > F
C. Total	24	62788.000		0.4318

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	239.74769	35.81946	6.69	<.0001*
Shelf Space	1.1438462	1.429352	0.80	0.4318

Figure 2: Plots of the residuals

Residual plot



Normal quantile plot of residuals.