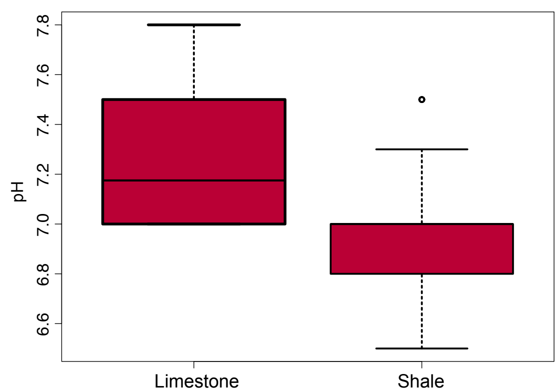
Total: 340 points

1. (15 pts) Consider the following box plots of stream pH on streams with limestone substrate and shale substrate.



a) Looking at these data, can you tell whether the t-test would be statistically significant? Why or why not. (7 pts)

b) Approximately what is the 97.5th percentile for the Shale measurements? (3 pts)

c) What does it mean that the boxplot for Limestone has no vertical line beneath it? (3 pts)

d) Why might a t-test be inappropriate for these data? (2 pts)

2. (10 pts) Define heteroskedacity in the context of linear regression and explain why heteroskedacity is a problem.

3. (5 pts) Let's say you have a two-way crossed ANOVA design. What change do you need to make to the following expression to account for an unbalanced design in which all treatment combinations are represented, but the number of replicates in each combination is not the same?

$$\sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{k=1}^{n} (Y_{ijk} - \bar{Y})^2$$

4. (3 pts each; 9 pts total) For each expression below, re-write the expression simplifying all sums possible and pulling outside of the sum any terms that can be moved outside of the sum.

   a. $\sum_{i=1}^{n} (a + bx_i + cy_i)$

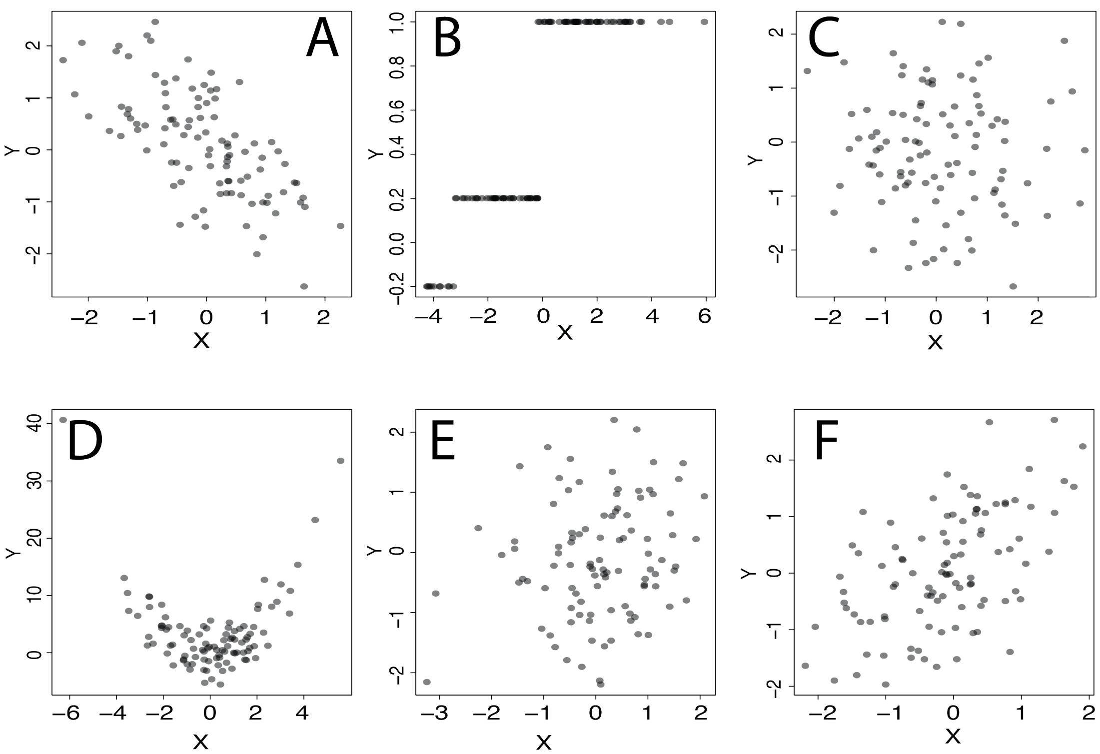
   b. $\sum_{i=1}^{n} \left(\frac{a}{e^{x_i}} \right)$

c. $\sum_{i=1}^{n}\sum_{j=1}^{m}(a + bx_i y_j + cz_j)$

5. (17 pts)

a) Rank the following plots in order of their Pearson's product moment correlation coefficient (regardless of whether Pearson's product moment correlation is appropriate) (6 pts):

_____ < _____ < _____ < _____ < _____ < _____



b) Which (if any) of these panels display data that would be inappropriate for Pearson's product moment correlation and why? (5 pts)

c) Name at least three aesthetic changes you would make to the figure in accordance with our discussion of "best practices" in Week 7 (6 pts).

6. (15 pts) Two researchers are pondering some data and they are trying to find the best distribution for that data. One researcher argues that the Gamma distribution is best:

$$f(x|\alpha,\beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}$$

whereas the other one argues that the Chi-squared distribution is best:

$$f(x \mid \nu = n) = \frac{1}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})} x^{(\frac{\nu}{2}-1)} e^{\frac{-x}{2}}$$

How should the researcher decide which one of them is correct? (Be as specific as possible! There are several possible ways to answer this question but only way that will give you full credit.)

7. (35 pts)

In the Week 13 problem set, you fit a model of the following form (I have written this in a slightly different format to be very clear about what I'm asking but the model is the same one you fit on the problem set' in this case I have re-named z as β_1 and I have renamed c as β_0):

$$S_i \sim Pois(\lambda_i)$$
$$log(\lambda_i) = \beta_0 + \beta_1 log(A_i) + \beta_2 Perimeter_i + \beta_3 Latitude_i + \beta_4 Longitude_i + \beta_5 SST_{j(i)} + \beta_6 Isolation_{j(i)}$$

where S_i was the number of species at a site i and SST and Isolation were measured for the island j on which site i sits.

a) How many parameters are associated with this model? (5 pts)

b) Why do we have to use the log link function for this kind of model? (10 pts)

Because of the relatively low spatial resolution of the environmental variables (e.g., SST), information on these covariates were drawn from a dataset that included information at the island level, but not at the site level.

c) Does this model represent an example of pseudoreplication, why or why not? (Full credit requires a definition of pseudoreplication.) (10 pts)

d) One suggestion for fixing the problem would be a hierarchical model of the following form:

$$S_i \sim Pois(\lambda_i)$$
$$log(\lambda_i) = \beta_0 + \beta_1 log(A_i) + \_\_\_\_\_\_\_$$

$$\beta_0 \sim N(\_\_, \_\_\_)$$

$$\beta_1 \sim N(\_\_, \_\_\_)$$

How would you include the covariates $Perimeter$, $Latitude$, $Longitude$, SST, and $Isolation$ in this new model to avoid potential problems of pseudoreplication? (10 pts)

8. (75 pts) Xiao et al. (2011) considered the use of nonlinear regression models for biological data.

Consider the following power-law model

$y = ax^b + \varepsilon, \;\; where\; \varepsilon \sim N(0, \sigma^2)$ [non-linear regression, which they call NLR]

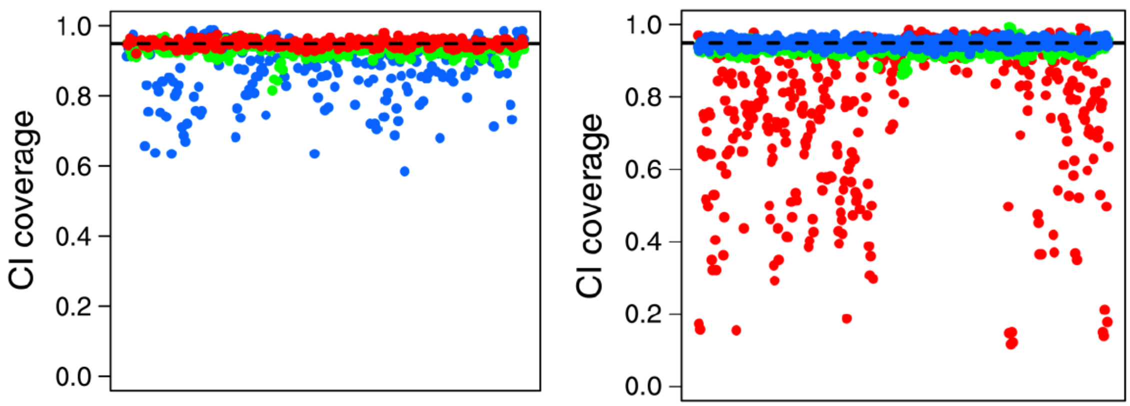
Some authors might choose to log both sides and fit a model of the form

$\log(y) = \log(a) + b log(x) + \varepsilon, \;\; where\; \varepsilon \sim N(0, \sigma^2)$ [linear regression, which they call LR]

a) Why are these models different? (5 pts)

In real situations, we do not know which of these models might be more appropriate, so we have to assume one form or the other when modelling our data. Xiao et al. (2011) simulated data using the NLR model and the LR model and then fit the NLR and LR models to those two datasets to see what happens when the incorrect model is used (and how it compares to when the correct model is used). They compare these model fits by simulating a large number of datasets (so the true values of the parameters are known) and looking at the estimates and confidence intervals relative to the known true value.

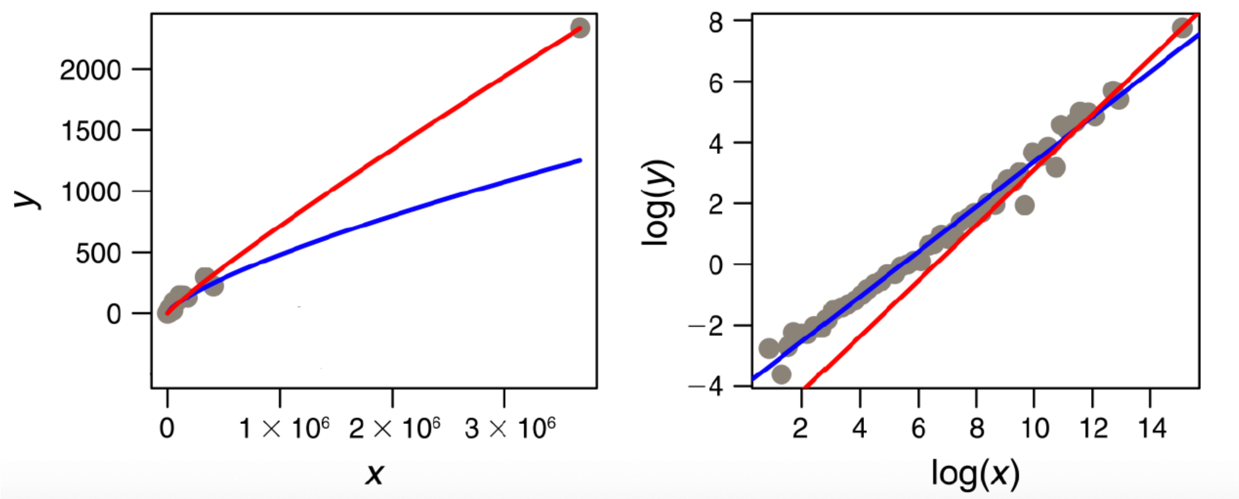
The figure below represents the confidence interval coverage for the b parameter when data are simulated according to the NLR (left panel) or LR (right panel) model and fit using the NLR model (red), the LR model (blue), or a weighted average of the two models (green). The horizontal dashed line represents 0.95.



b) Define the “coverage” of a confidence interval. (6 pts)

c) Why is it a problem for hypothesis testing if the coverage of your confidence intervals is too small? (6 pts)

d) Consider the empirical data below, fit using the two models and plotted on either the x,y scale or the log(x),log(y) scale alongside the LR fit (blue) and NLR fit (red). (Each plot shows the very same data and the same lines, just plotted differently.) As a modeler, what concerns might I have about modelling these data on the x,y scale illustrated at left? (10 pts)



e) Why are the LR and NLR models so different on the x,y scale illustrated at left? (Hint: This is related to the previous question.) (10 pts)

f) Because we don't know which model is more appropriate for real datasets, we might want to compare these models. Why would a likelihood ratio test (LRT) be inappropriate here? (10 pts)

g) Because the LRT is inappropriate, the authors use AIC instead. What is the formula for AIC? (6 pts)

The authors suggest that in cases where there is no biological knowledge about the correct model, model averaging should be used. They cite a study of eyeball size and body mass size (the subject of an earlier question on this exam) that found

$$AIC_{LR} - AIC_{NLR} = 7.9$$

h) Given the above information, the data provide more support for which of these two models? (6 pts)

i) What is the expression for the model weight for the two models? (6pts)

j) Cycling back to the plots of coverage for the two models (after part a) and the AIC weighted model, why does it make sense that the AIC weighted model would perform better overall than the two models fit individually? (10 pts)

9. (55 pts)
This question will focus on the analysis completed for the following paper:

RESEARCH ARTICLE

Hangry in the field: An experience sampling study on the impact of hunger on anger, irritability, and affect

Viren Swami[1,2], Samantha Hochstöger[3], Erik Kargl[3], Stefan Stieger[3]*

1 School of Psychology and Sport Science, Anglia Ruskin University, Cambridge, United Kingdom, 2 Centre for Psychological Medicine, Perdana University, Kuala Lumpur, Malaysia, 3 Department of Psychology and Psychodynamics, Karl Landsteiner University of Health Sciences, Krems an der Donau, Austria

The basic idea behind the study can be captured by the abstract:

Abstract

The colloquial term "hangry" refers to the notion that people become angry when hungry, but very little research has directly determined the extent to which the relationship between hunger and negative emotions is robust. Here, we examined associations between every-day experiences of hunger and negative emotions using an experience sampling method. Sixty-four participants from Central Europe completed a 21-day experience sampling phase in which they reported their hunger, anger, irritability, pleasure, and arousal at five time-points each day (total = 9,142 responses). Results indicated that greater levels of self-reported hunger were associated with greater feelings of anger and irritability, and with lower pleasure. These findings remained significant after accounting for participant sex, age, body mass index, dietary behaviours, and trait anger. In contrast, associations with arousal were not significant. These results provide evidence that everyday levels of hunger are associated with negative emotionality and supports the notion of being "hangry".

Though the original study looked at a suite of responses, this question will focus on just "Irritability". The authors considered 6 model covariates that might explain Irritability: **Sex**: Binary (male=0 or female=1), **Age**: Continuous, **BMI (Body Mass Index)**: Continuous, **DB (Dietary Behavior)**: Discrete (4 categories: restrictive, driven by emotion, driven by boredom, driven by external influences like friends and family), **BPAQ (Buss and Perry Aggression Questionnaire)**: continuous but bounded between 0 (no anger) and 1 (very angry), **Hunger (Time since last meal)**: Continuous and non-negative. Assume throughout the questions that follow that Age, BMI, BPAQ are standardized such that after standardization, the mean and standard deviation of these covariates is 0 and 1, respectively.

a) Assuming "Irritability" follows a Gaussian distribution, write out the full model equation for a model including all of these covariates. (15 pts)

b) How many parameters does this model have? (5 pts)

c) In words, how would we interpret the intercept of this model? (5 pts)

d) The authors also collected additional demographic information from participants, including nationality (Austria, Germany, Switzerland, etc.). Do you think the authors should model nationality as a random effect or a fixed effect and why? (10 pts)

e) Assuming the authors followed your advice in part d, how many additional parameters does it add to the model to include nationality as a covariate? (5 pts)

f) Describe two methods by which the authors could assess whether nationality should be included in the final model or not. (15 pts)

10. (70 pts)

Schneider et al. (2017) designed a study to assess whether the anticipation of pride or the anticipation of guilt had more influence (and in what direction) on an environmentally conscious decision. A sample of 545 U.S. participants were presented with two decisions in which they could choose an environmentally-friendly choice (the “green” choice; example: installing geothermal) or one that was less environmentally friendly (the “brown” choice; example: using natural gas heating), and each participant was also asked a series of questions to evaluate their emotional state of which we will focus on just two: feelings of pride and feeling of guilt. These two feeling were each measured on what is known as a Likert-scale which is a discrete scale from 1 to 5. Feelings of pride ranged from 1 (no pride) to 5 (full of pride) and feeling of guilt were measured from 1 (no guilt) to 5 (guilt). Note that pride and guilt are separate variables, so an increase in one does not imply a decrease in the other.

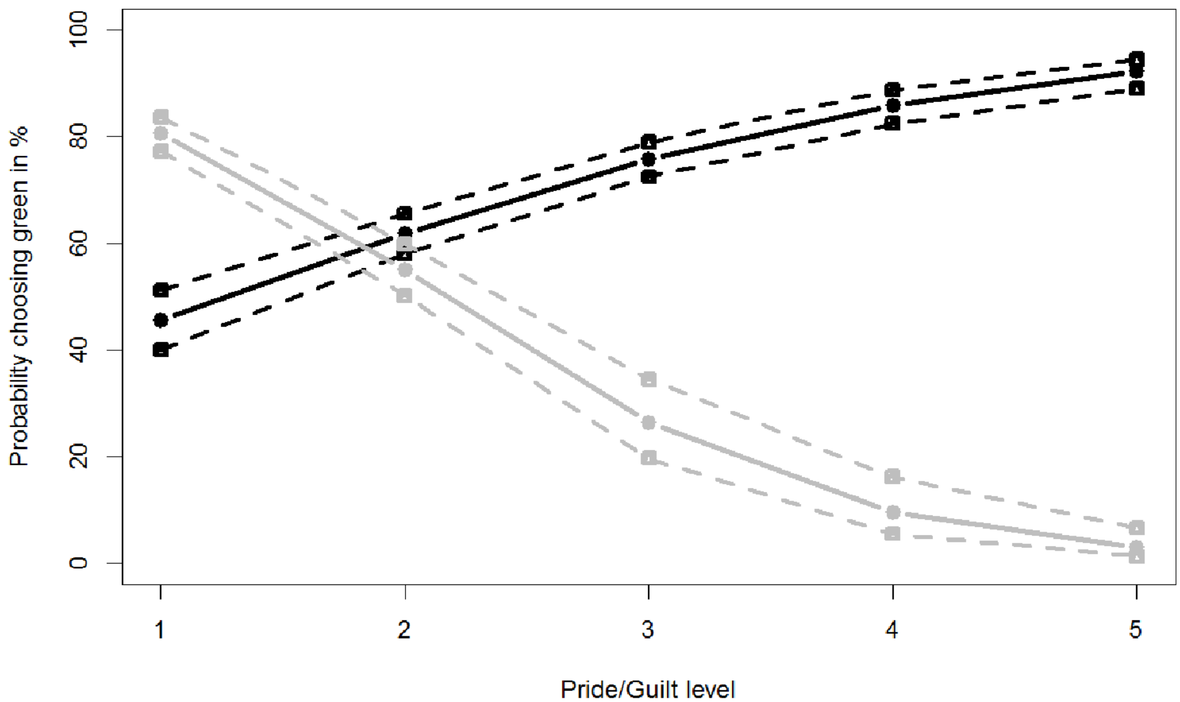
First, we will consider only the independent variable representing feeling of pride.

a) Ignoring guilt for a moment, write the mathematical expression(s) that describe the model that the authors should use to test whether feelings of pride influence the probability of making a pro-environmental decision. (10 pts)

b) Write out the joint likelihood for this model. (5 pts)

c) The authors find that the coefficient for the variable pride (measured on the Likert scale) is 0.66 (SE=0.06). In words, how would you interpret this number in terms of the probability of making a pro-environmental decision (be as specific as possible). (5 pts)

d) The authors fit a separate model for the variable "pride" and the variable "guilt", the results of which are plotted below (pride results are shown in black; guilt results are shown in gray). The confidence intervals for the guilt curve are narrowest at the ends and widest in the middle, but for the pride curve they are widest on the left hand side of the plot – why? (10 pts)



e) In the actual study, the participants then considered models with additional covariates (Gender, Age, Education, Political Party, Income, and "Care for the Environment"). Such a model might suffer from multi-collinearity. What impact would such collinearity have on the model results and how might be researchers change their model to accommodate? (15 pts)

f) Lets say that the authors wanted to quickly find more study participants and so they ask each randomly selected participant to give the same questionnaire to three friends. Does this simply quadruple the study size or do the modelers need to make any changes to their design – why or why not? (10 pts)

g) Define statistical power and then name two different ways that the authors could increase their statistical power in this study. (15 pts)

11. (34 pts) The study described in the previous question was actually just a pilot study to the main goal, which was to understand what kinds of prompts might change environmental decision making. This case, the authors did a 3 x 2 crossed study design where they used three types of prompts[1] and those prompts could be geared towards two different feelings (guilt or pride). We will assume that both prompt type and guilt/pride are modelled as fixed effects and we want to consider their interaction. After one of these 6 possible combinations (3 prompt types crossed with two emotions), respondents were asked to make 5 environmental decisions in which they rated their likelihood of buying a product on a scale of 1 (*not at all likely*) to 7 (*extremely likely*). We will assume for a moment that an ANOVA is the appropriate approach to understanding whether different prompts yield different consumer choices.

Fill in all blank cells of the ANOVA table below (21 pts)

Source	SS	dof	MS (leave as ratio)	F-ratio (leave as ratio)	p-value
					See below
					Leave blank
					Leave blank
				N/A	N/A
Total			N/A	N/A	N/A

[1] the following details are not important to the question, but the prompts were: 1 – a one sentence reminder; 2 – affective forecasting, where they were asked to predict whether a decision would make them feel pride or guilt; 3 – a writing prompt in which they wrote an essay on the implications of their future choice

What is the R command to calculate the p-value for the cell indicated above? (5 pts)

In words, how would we interpret the interaction term in this scenario? (8 pts)

Two sample t-test (equal variances)	$\frac{(\bar{X}_A-\bar{X}_B)-(\mu_A-\mu_B)}{\sqrt{s^2_{pooled}\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}} \sim t_{n_A+n_B-2}$ $s^2_{pooled}=\frac{1}{n_A+n_B-2}(SS_A+SS_B)$	$H_0: \mu_A=\mu_B$	$\frac{(\bar{X}_A-\bar{X}_B)}{\sqrt{s^2_{pooled}\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}} \sim t_{n_A+n_B-2}$ $s^2_{pooled}=\frac{1}{n_A+n_B-2}(SS_A+SS_B)$	$T=\frac{(\bar{X}_A-\bar{X}_B)}{\sqrt{s^2_{pooled}\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}}$	$(\bar{X}_A-\bar{X}_B)+\sqrt{s^2_{pooled}\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}t_{(\frac{\alpha}{2})[n_A+n_B-2]} \le \mu_A-\mu_B \le (\bar{X}_A-\bar{X}_B)+\sqrt{s^2_{pooled}\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}t_{(1-\frac{\alpha}{2})[n_A+n_B-2]}$ $(\bar{X}_A-\bar{X}_B)-\sqrt{s^2_{pooled}\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}t_{(1-\frac{\alpha}{2})[n_A+n_B-2]} \le \mu_A-\mu_B \le (\bar{X}_A-\bar{X}_B)+\sqrt{s^2_{pooled}\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}t_{(1-\frac{\alpha}{2})[n_A+n_B-2]}$
	$\frac{(n-1)s^2}{\sigma^2}\sim\chi^2_{n-1}$	We didn't cover hypothesis testing on variances			$\frac{(n-1)s^2}{\chi^2_{(1-\frac{\alpha}{2})[n-1]}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{(\frac{\alpha}{2})[n-1]}}$
F-test	$\frac{\sigma_A^2/\sigma_B^2}{s_A^2/s_B^2}\sim F_{m-1,n-1}$	$H_0: \sigma_A^2=\sigma_B^2$	$\frac{s_A^2}{s_B^2}\sim F_{n-1,m-1}$ *Note the order of the parameters	$T=\frac{s_A^2}{s_B^2}$	$\frac{s_A^2}{s_B^2}F_{(\frac{\alpha}{2})[m-1,n-1]} \le \sigma_A^2/\sigma_B^2 \le \frac{s_A^2}{s_B^2}F_{(1-\frac{\alpha}{2})[m-1,n-1]}$
Wald test for binomial proportions	$\frac{\hat{p}-\theta}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}\sim N(0,1)$	$H_0: \theta=\theta_0$	$\frac{\hat{p}-\theta_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}\sim N(0,1)$	$T=\frac{\hat{p}-\theta_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$	$\hat{p}-z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le \theta \le \hat{p}+z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $\hat{p}+z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le \theta \le \hat{p}+z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
Score test for binomial proportions	$\frac{\hat{p}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}}\sim N(0,1)$	$H_0: \theta=\theta_0$	$\frac{\hat{p}-\theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}}\sim N(0,1)$	$T=\frac{\hat{p}-\theta_0}{\sqrt{\frac{\theta_0(1-\theta_0)}{n}}}$	Not easily inverted to form a confidence interval

	$T\sim f(T)$	H_0	$T\vert H_0\sim f(T\vert H_0)$	$T\vert H_0$	**Confidence interval on parameter of interest**
One-sample t-test; σ^2 known	$\frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}}\sim N(0,1)$	$H_0: \mu=c$	$\frac{\bar{X}-c}{\sqrt{\frac{\sigma^2}{n}}}\sim N(0,1)$	$T=\frac{\bar{X}-c}{\sqrt{\frac{\sigma^2}{n}}}$	$\bar{X}-\sqrt{\frac{\sigma^2}{n}}z_{1-\alpha/2} \le \mu \le \bar{X}+\sqrt{\frac{\sigma^2}{n}}z_{1-\alpha/2}$ $\bar{X}+\sqrt{\frac{\sigma^2}{n}}z_{\alpha/2} \le \mu \le \bar{X}+\sqrt{\frac{\sigma^2}{n}}z_{1-\alpha/2}$
One-sample t-test; σ^2 unknown	$\frac{\bar{X}-\mu}{\sqrt{\frac{s^2}{n}}}\sim t_{n-1}$	$H_0: \mu=c$	$\frac{(\bar{X}-c)}{\sqrt{\frac{s^2}{n}}}\sim t_{n-1}$	$T=\frac{(\bar{X}-c)}{\sqrt{\frac{s^2}{n}}}$	$\bar{X}-\sqrt{\frac{s^2}{n}}t_{(1-\alpha/2)[n-1]} \le \mu \le \bar{X}+\sqrt{\frac{s^2}{n}}t_{(1-\alpha/2)[n-1]}$ $\bar{X}+\sqrt{\frac{s^2}{n}}t_{(\alpha/2)[n-1]} \le \mu \le \bar{X}+\sqrt{\frac{s^2}{n}}t_{(1-\alpha/2)[n-1]}$
Two sample t-test (unequal variances)	$\frac{(\bar{X}_A-\bar{X}_B)-(\mu_A-\mu_B)}{\sqrt{\frac{s_A^2}{n_A}+\frac{s_B^2}{n_B}}}\sim t_{dof}$ $dof=\frac{\left[\frac{s_A^2}{n_A}+\frac{s_B^2}{n_B}\right]^2}{\frac{\left(\frac{s_A^2}{n_A}\right)^2}{n_A-1}+\frac{\left(\frac{s_B^2}{n_B}\right)^2}{n_B-1}}$	$H_0: \mu_A=\mu_B$	$\frac{(\bar{X}_A-\bar{X}_B)}{\sqrt{\frac{s_A^2}{n_A}+\frac{s_B^2}{n_B}}}\sim t_{dof}$ $dof=\frac{\left[\frac{s_A^2}{n_A}+\frac{s_B^2}{n_B}\right]^2}{\frac{\left(\frac{s_A^2}{n_A}\right)^2}{n_A-1}+\frac{\left(\frac{s_B^2}{n_B}\right)^2}{n_B-1}}$	$T=\frac{(\bar{X}_A-\bar{X}_B)}{\sqrt{\frac{s_A^2}{n_A}+\frac{s_B^2}{n_B}}}$	$(\bar{X}_A-\bar{X}_B)-\sqrt{\frac{s_A^2}{n_A}+\frac{s_B^2}{n_B}}t_{(1-\frac{\alpha}{2})[dof]} \le \mu_A-\mu_B \le (\bar{X}_A-\bar{X}_B)+\sqrt{\frac{s_A^2}{n_A}+\frac{s_B^2}{n_B}}t_{(1-\frac{\alpha}{2})[dof]}$ $(\bar{X}_A-\bar{X}_B)+\sqrt{\frac{s_A^2}{n_A}+\frac{s_B^2}{n_B}}t_{(\frac{\alpha}{2})[dof]} \le \mu_A-\mu_B \le (\bar{X}_A-\bar{X}_B)+\sqrt{\frac{s_A^2}{n_A}+\frac{s_B^2}{n_B}}t_{(1-\frac{\alpha}{2})[dof]}$